

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
7. November 2002 (07.11.2002)

PCT

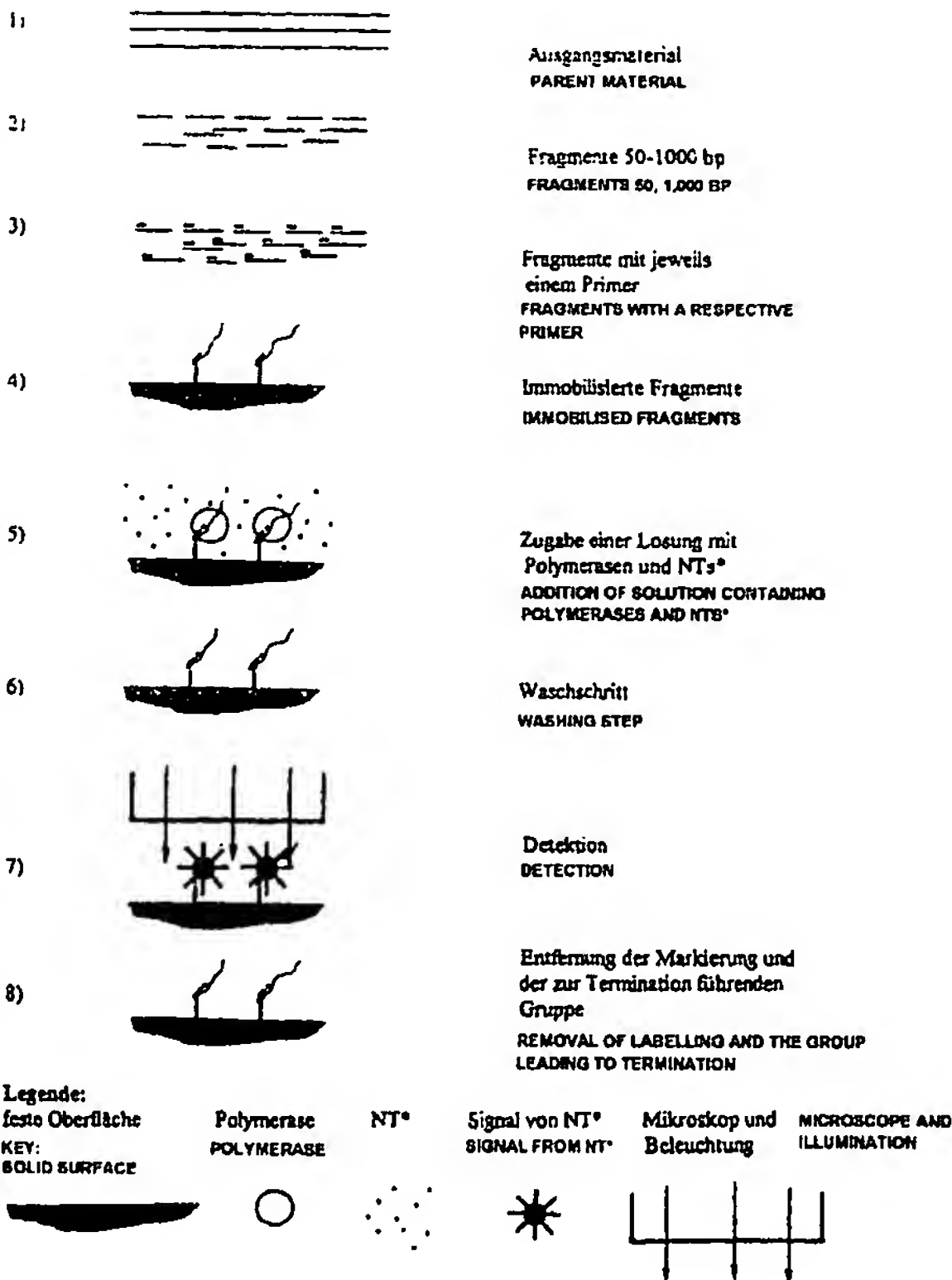
(10) Internationale Veröffentlichungsnummer  
WO 02/088382 A2

- (51) Internationale Patentklassifikation<sup>7</sup>: C12Q 1/00 (72) Erfinder; und  
(21) Internationales Aktenzeichen: PCT/EP02/04659 (75) Erfinder/Anmelder (nur für US): TCHERKASSOV, Dmitri [RU/DE]; Kahlhorststrasse 36, 23562 Lübeck (DE).  
(22) Internationales Anmeldedatum: 26. April 2002 (26.04.2002) (74) Anwälte: WEBER-QUITZAU, Martin usw.; Uexküll & Stolberg, Beselerstrasse 4, 22607 Hamburg (DE).  
(25) Einreichungssprache: Deutsch (81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.  
(26) Veröffentlichungssprache: Deutsch  
(30) Angaben zur Priorität: 101 20 797.2 27. April 2001 (27.04.2001) DE  
(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): GENOVOXX GMBH [DE/DE]; Ratzeburger Allee 160, 23538 Lübeck (DE).

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR ANALYSING NUCLEIC ACID CHAINS

(54) Bezeichnung: VERFAHREN ZUR ANALYSE VON NUKLEINSÄUREKETTEN



(57) Abstract: The invention relates to a method for analysing nucleic acid chains. The basis for the method is the detection of the fluorescent signals of individual nucleotide molecules, labelled with dyes, which are incorporated into growing nucleic acid chains by a polymerase. The reaction takes place on a planar surface. A large number of individual nucleic acid molecules are immobilised on said surface. All these nucleic acid molecules are subjected to the same conditions, so that a synthesis reaction can take place simultaneously in all the nucleic acid molecules.

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Analyse von Nukleinsäureketten. Grundlage der Methode ist die Detektion von Fluoreszenzsignalen einzelner, mit Farbstoffen markierter Nukleotidmoleküle, die durch eine Polymerase in wachsende Nukleinsäureketten eingebaut werden. Die Reaktion verläuft auf einer planen Oberfläche. Auf dieser Oberfläche sind viele einzelne Nukleinsäure-Moleküle immobilisiert. Alle diese Nukleinsäure-Moleküle sind gleichen Bedingungen ausgesetzt, so dass an allen Nukleinsäure-Molekülen gleichzeitig eine Aufbaureaktion ablaufen kann.

WO 02/088382 A2

BEST AVAILABLE COPY



(84) Bestimmungsstaaten (*regional*): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Veröffentlicht:**

— ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

### Verfahren zur Analyse von Nukleinsäureketten

Die Erfindung betrifft ein Verfahren zur Analyse von Nukleinsäureketten. Die Grundlage der Methode ist die Detektion von Fluoreszenzsignalen einzelner mit Farbstoffen markierter Nukleotidmoleküle, die durch eine Polymerase in eine wachsende Nukleinsäurekette eingebaut werden. Die Reaktion verläuft auf einer planen Oberfläche. An diese Oberfläche sind viele einzelne Nukleinsäure-Moleküle gebunden. Alle diese Nukleinsäure-Moleküle sind gleichen Bedingungen ausgesetzt, so dass an allen Nukleinsäure-Molekülen gleichzeitig eine Aufbaureaktion ablaufen kann.

- 2 -

Das Verfahren umfaßt im wesentlichen folgende Schritte:

- 1) Bindung der Nukleinsäureketten (NSKFs) auf einer planen Oberfläche mit anschließender Hybridisierung von Primern, alternativ Bindung von Primern mit anschließender Hybridisierung von NSKFs, so dass NSKF-Primer-Komplexe gebildet werden.
  - 2) Durchführen einer zyklischen Aufbaureaktion, wobei jeder Zyklus aus folgenden Schritten besteht:
    - a) Zugabe einer Lösung mit markierten Nukleotiden (NTs') und Polymerase zu den gebundenen NSKF-Primer-Komplexen,
    - b) Inkubation der gebundenen NSKF-Primer-Komplexe mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
    - c) Waschen,
    - d) Detektion der Signale von einzelnen Molekülen,
    - e) Entfernung der Markierung von den eingebauten Nukleotiden,
    - f) Waschen.
- Gegebenenfalls erfolgen mehrfache Wiederholungen des Zyklus.
- 3) Analyse der detektierten Signale der einzelnen Moleküle.
  - 4) Rekonstruktion der Sequenzen aus den Einzeldaten.

#### 1. Abkürzungen und Begriffserläuterungen

DNA - Desoxyribonukleinsäure verschiedenen Ursprungs und unterschiedlicher Länge (genomische DNA, cDNA, ssDNA, dsDNA)

RNA - Ribonukleinsäure (meist mRNA)

Polymerasen - Enzyme, die komplementäre Nukleotide in einen wachsenden DNA- oder RNA-Strang einbauen können (z.B. DNA-Polymerasen, Reverse-Transkriptasen, RNA-Polymerasen)

dNTP - 2'-deoxy-Nucleosid-Triphosphate, Substrate für DNA-



- 3 -

## Polymerasen und Reverse-Transkriptasen

NTP - Nukleosid-Triphosphate, Substrate für RNA-Polymerasen

5 NT - natürliches Nukleotid, meist dNTP, wenn nicht ausdrücklich anders gekennzeichnet.

Abkürzung "NT" wird auch bei der Längenangabe einer Nukleinsäuresequenz verwendet, z.B. 1.000 NT. In diesem Fall steht "NT"  
10 für Nukleosid-Monophosphate.

Im Text wird bei Abkürzungen die Mehrzahl durch Verwendung des Suffixes "s" gebildet, "NT" steht zum Beispiel für "Nukleotid", "NTs" steht für mehrere Nukleotide.

15

NT' - modifiziertes Nukleotid, meist dNTP, wenn nicht ausdrücklich anders gekennzeichnet. NTs' bedeutet: modifizierte Nukleotide

20 NSK - Nukleinsäurekette. DNA oder RNA in ihrer ursprünglichen Länge

NSKF - Nukleinsäurekettenfragment (DNA oder RNA), das einem Teil der Gesamtsequenz entspricht, NSKFs -  
25 Nukleinsäurekettenfragmente. Die Summe der NSKFs bildet ein Äquivalent zur Gesamtsequenz. Die NSKFs können beispielsweise Fragmente von DNA- oder RNA-Gesamtsequenz sein, die nach einem Fragmentierungsschritt entstehen.

30 Gesamtsequenz - die in der Sequenzierungsreaktion eingesetzte Sequenz oder die eingesetzten Sequenzen, meistens in NSKFs überführt. Sie kann ursprünglich aus einer oder mehreren NSKs bestehen. Dabei kann die Gesamtsequenz Teile oder Äquivalente einer anderen Sequenz oder von Sequenz-Populationen darstellen  
35 (z.B. mRNA, cDNA, Plasmid-DNA mit Insert, BAC, YAC) und aus einer oder unterschiedlichen Spezies stammen.

Primerbindungstelle (PBS) - Teil der Sequenz in der NSK oder

- 4 -

NSKF, an den der Primer bindet.

Referenzsequenz - eine bereits bekannte Sequenz, zu der die Abweichungen in der zu untersuchenden Sequenz bzw. in den zu  
5 untersuchenden Sequenzen (Gesamtsequenz) ermittelt werden. Als Referenzsequenzen können in Datenbanken zugängliche Sequenzen verwendet werden, wie z.B. aus der NCBI-Datenbank.

T<sub>m</sub> - Schmelztemperatur

10

Plane Oberfläche: Oberfläche, die vorzugsweise folgende Merkmale aufweist: 1) Sie erlaubt, mehrere einzelne Moleküle, vorzugsweise mehr als 100, noch besser mehr als 1000, mit dem jeweiligen gegebenen Objektiv-Oberfläche-Abstand bei einer  
15 Objektivposition gleichzeitig zu detektieren. 2) Die immobilisierten einzelnen Moleküle befinden sich in derselben Fokusebene, die reproduzierbar eingestellt werden kann.

Weitfeld-Optik-Detektionssystem - Detektionssystem, das  
20 gleichzeitig Fluoreszenzsignale von einzelnen, auf einer Fläche verteilten Molekülen detektieren kann, wobei die Fläche ca. 100  $\mu\text{m}^2$  und größer ist. Ein Beispiel für Weitfeld-Detektionsoptik stellt Fluoreszenzmikroskop Axiovert 200 oder Axioplan 2e (Zeiss) mit einem Planneofluar-Objektiv 100x NA  
25 1.4 Ölimmersion (Zeiss), oder einem Planapochromat-Objektiv 100x NA 1.4 Ölimmersion (Zeiss); die Anregung der Fluoreszenz kann dabei mit einer Lampe, z.B. Quecksilberdampf Lampe, oder einem Laser oder Dioden erfolgen. Sowohl Epifluoreszenzmodus als auch im Totalreflexions-Fluoreszenzmikroskopie-Modus  
30 (total internal reflection fluorescence microscopy, TIRF-Microscopy) oder Laser-Scanning-Mikroskopie-Modus können verwendet werden. In dieser Anmeldung wird Gebrauch von dieser Weitfeld-Detektionsoptik gemacht

35

Sterisches Hindernis: Sterisch anspruchsvolle Gruppe, die durch ihre chemische Struktur die Eigenschaften der mit dieser Gruppe gekoppelten NTs' so verändert, dass diese durch eine Polymerase

- 5 -

in einer Extensionsreaktion nicht nacheinander eingebaut werden können.

Definition der Termination: Als Termination wird in dieser Anmeldung der reversible Stop des Einbaus der modifizierten ungespalteten NTs bezeichnet.

Dieser Begriff ist von dem üblichen Gebrauch des Wortes "Termination" durch Dideoxy-NTP bei einer konventionellen Sequenzierung zu trennen.

10

Die Termination kommt nach dem Einbau eines modifizierten NT zustande. Das modifizierte eingebaute NT trägt eine an die Base reversibel gekoppelte sterische Gruppe, die zur Behinderung des Einbaus eines nächsten komplementären NT in den wachsenden Strang durch eine Polymerase führt.

Genprodukte - Bei den Genprodukten handelt es sich um die primären Genprodukte der Gene. Im wesentlichen handelt es sich dabei um RNA-Transkripte der genannten Gene, welche auch als Target-Sequenzen (oder Target-Nukleinsäuresequenzen) bezeichnet werden. Diese Target-Sequenzen schließen neben mRNA auch davon abgeleitete einzelsträngige und doppelsträngige cDNA, von cDNA abgeleitete RNA oder von cDNA amplifizierte DNA ein.

25 Einzelnukleotidpolymorphismen (single nucleotide polymorphisms, SNPs) - Veränderungen in den Sequenzen, die als Substitution (Transition oder Transversion) oder als Deletion oder Insertion einzelner NT auftreten können.

30

## 2. Stand der Technik

Die Nukleinsäurenketten-Sequenzanalyse ist in vielen Bereichen der Wissenschaft, Medizin und Industrie zu einem wichtigen Werkzeug geworden. Zur Analyse wurden mehrere Verfahren entwickelt.

Die bekanntesten Verfahren sind die Ketten-Terminations-Sequen-

- 6 -

zierung nach Sanger (F. Sanger et al. PNAS 1977 v.74 s. 5463), die auf dem Einbau von Kettenterminatoren basiert, und die Maxam-Gilbert-Methode, die auf Basen-spezifischer Modifikation und Spaltung von Nukleinsäureketten beruht (A.M. Maxam and W. 5 Gilbert PNAS 1977, v.74 S.560). Beide Methoden liefern eine Anzahl von Nukleinsäurekettenfragmenten verschiedener Längen. Diese Fragmente werden der Länge nach in einem Gel aufgetrennt. Dabei müssen alle Nachteile der Elektrophorese (wie z.B. lange Laufzeit, relativ kurze Strecken von Sequenzen, die in einem 10 Ansatz bestimmt werden können, begrenzte Anzahl der parallelen Ansätze sowie relativ große Mengen an DNA) in Kauf genommen werden. Diese Methoden sind sehr arbeitsintensiv und langsam.

Ein weiteres Verfahren zur Sequenzierung basiert auf der Hybri- 15 disierung von Nukleinsäureketten mit kurzen Oligonukleotiden. Dabei wird mit mathematischen Methoden berechnet, wie viele Oligonukleotide einer bestimmten Länge vorhanden sein müssen, um eine komplette Sequenz zu ermitteln (Z.T. Strezoska et al. PNAS 1991 v.88 S.10089, R.S.Drmanac et al. Science 1993 v.260 20 S.1649). Auch dieses Verfahren ist mit Problemen behaftet: Es kann nur eine Sequenz in einem Ansatz bestimmt werden, sekundäre Strukturen stören die Hybridisierung und Sequenzwiederholungen verhindern die korrekte Analyse.

25 Eine andere Möglichkeit zur Sequenzierung haben Arbeitsgruppen beispielsweise von (Dower US Patent 5.547.839, Canard et al. US Patent 5.798.210, Rasolonjatovo Nucleosides & Nucleotides 1999, v.18 S.1021, Metzker et al. NAR 1994, v.22, S.4259, Welch et al. Nucleosides & Nucleotides 1999, v.18, S.197) entwickelt. 30 Diese Methode wird abgekürzt als BASS (Base Addition Sequencing Scheme) oder SBS (Sequencing by Synthesis) bezeichnet. Dabei wird eine große Anzahl gleicher einzelsträngiger DNA-Stücke an einem definierten Ort auf einer Oberfläche fixiert und das Signal von der Gesamtheit dieser vielen identischen DNA-Stücke 35 analysiert. Zu dieser fixierten DNA wird eine Lösung mit Polymerase und Nukleotiden zugegeben, so dass ein komplementärer Strang synthetisiert werden kann. Dabei soll die Polymerase schrittweise arbeiten: in jedem Schritt wird nur ein

- 7 -

einziges Nukleotid eingebaut. Dieses wird detektiert, worauf die Polymerase in einem nächsten Zyklus das nächste Nukleotid einbaut. Bei dieser Methode wurden an der 3'-OH-Gruppe der Deoxyribose modifizierte Nukleotide verwendet.

5 Trotz des Gelingens einiger einzelner Schritte der Methode wurde sie nicht zu einem funktionsfähigen Verfahren entwickelt. Dies kann beispielsweise auf folgenden Tatsachen beruhen: Beim Aufbau der komplementären Stränge tritt sehr schnell eine Desynchronisation der Synthese auf, so dass bei  
10 jedem Schritt die Fehler akkumulieren. Deshalb können nur sehr kurze Fragmente sequenziert werden. Es ist zu betonen, dass alle beschriebenen BASS-Methoden nicht auf der Detektion von einzelnen Molekülen beruhen. Das Signal wird stattdessen von einer großen Anzahl identischer an einem definierten Ort  
15 immobilisierter Moleküle registriert. Die in diesen Methoden übliche Verwendung der Begriffe "einzelne Moleküle" und "Moleküle" zielt dabei nicht auf individuelle, voneinander getrennte Moleküle, sondern auf eine Population, die aus vielen identischen Molekülen besteht. Identisch heißt in diesem Fall,  
20 dass die Moleküle die gleiche Sequenz haben.

Ein weiteres Problem stellen die an 3'-Position modifizierten Nukleotide dar. Einerseits führt die Ankopplung großer Substituenten an diese Position der Nukleotide in vielen Fällen zur Störung der Einbaureaktion durch die Polymerasen.  
25 Andererseits sind Polymerasen im Stande, einige Modifikationen von der 3'-Position abzuspalten.

Die Aufgabe der vorliegenden Erfindung besteht daher darin, ein Verfahren zur Sequenzanalyse von Nukleinsäureketten  
30 bereitzustellen, das die Nachteile der oben erwähnten Methoden nicht aufweist und vor allem eine billigere, schnellere und effizientere Analyse von Nukleinsäuresequenzen ermöglicht. Insbesondere soll das Verfahren in der Lage sein, viele Sequenzen parallel zu bestimmen. Es kann dann beispielsweise  
35 für die Analyse sehr langer Nukleinsäureketten (mehrere Mb) oder für die Variantenanalyse an vielen kurzen Ketten (Mutationsanalyse, SNP-Analyse) in einem Ansatz verwendet werden.



### 3. Kurze Beschreibung

5 Die Aufgabe wird erfindungsgemäß durch ein Verfahren zur parallelen Sequenzanalyse von Nukleinsäuresequenzen (Nukleinsäureketten, NSKs) gelöst, bei dem man

10 Fragmente (NSKFs) einzelsträngiger NSKs mit einer Länge von etwa 50 bis 1000 Nukleotiden erzeugt, die überlappende Teilsequenzen der Gesamtsequenzen darstellen, man

15 die NSKFs unter Verwendung eines einheitlichen oder mehrerer unterschiedlicher Primer in Form von NSKF-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

20 eine zyklische Aufbaureaktion des komplementären Stranges der NSKFs unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

a) zu den an die Oberfläche gebundenen NSKF-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs') enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs' jeweils an den NTs' befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs' durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs' strukturell an der Base so modifiziert sind, dass die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT' in denselben Strang einzubauen, wobei der Fluoreszenzfarbstoff abspaltbar ist und die strukturelle Modifikation ein abspaltbarer sterisch anspruchsvoller Ligand ist, man

25

30

35

- 9 -

- 5           b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT' verlängert werden, man
- 10           c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs' geeignet sind, man
- 15           d) die einzelnen, in komplementäre Stränge eingebauten NTs' durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- 20           e) zur Erzeugung unmarkierter (NTs oder) NSKFs die Fluoreszenzfarbstoffe und die sterisch anspruchsvollen Liganden von den am komplementären Strang angefügten NTs' abspaltet, man
- 25           f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Liganden geeignet sind, man

30           die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

wobei man die relative Position einzelner NSKF-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser NSKFs durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt.

Aus den ermittelten Teilsequenzen kann man beispielsweise die Gesamtsequenz der NSKs bestimmen. Unter einer parallelen



- 10 -

Sequenzanalyse wird in diesem Zusammenhang die gleichzeitige Sequenzanalyse vieler NSKFs verstanden (beispielsweise 1.000.000 bis 10.000.000), wobei diese NSKFs von einer einheitlichen NSK-Population oder von mehreren 5 unterschiedlichen NSK-Populationen abgeleitet sind.

Die erhaltene Population von überlappenden Teilsequenzen läßt sich beispielsweise bei de novo Sequenzierung mit kommerziell erhältlichen Programmen zur Gesamtsequenz der NSK zusammenfügen 10 (Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257).

Bei der Analyse von Varianten einer bekannten Referenzsequenz 15 lassen sich Mutationen oder Einzelnukleotidpolymorphismen durch einen Vergleich der erhaltenen überlappenden Teilsequenzen mit der Referenzsequenz feststellen.

Gemäß einer besonderen Ausführungsform der Erfindung kann das 20 Verfahren durchgeführt werden, indem man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man

- a) in jedem Zyklus nur jeweils ein markiertes NT',
- b) in jedem Zyklus jeweils zwei unterschiedlich markierte 25 NTs' oder
- c) in jedem Zyklus jeweils vier unterschiedlich markierte NTs'

einsetzt.

30

Wenn die NSKs Varianten einer bekannten Referenzsequenz sind kann das Verfahren auch durchgeführt werden, indem man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in den Zyklen abwechselnd jeweils zwei 35 unterschiedlich markierte NTs' und zwei unmarkierte NTs einsetzt und man die Gesamtsequenzen durch Vergleich mit der Referenzsequenz ermittelt.

- 11 -

Gegenstand der vorliegenden Erfindung sind ferner die in den Figuren 7e, 7f und 7g dargestellten Nukleotide und die entsprechenden markierten Nukleotide, die beispielsweise an die terminale Aminofunktion angeheftete Fluoreszenzfarbstoffe 5 aufweisen, oder die in den Figuren 7h, 7i oder 7j dargestellten markierten Nukleotide.

Gegenstand der vorliegenden Erfindung ist ferner die Verwendung der in Figuren 7e, 7f und 7g dargestellten Nukleotide und der 10 entsprechenden mit einem Fluoreszenzfarbstoff markierten Nukleotide für das erfindungsgemäßen Verfahren.

Gegenstand der vorliegenden Erfindung ist ferner die Verwendung der an der Base modifizierten NT\*s (Beispiele siehe Figuren 15 7k, 7L und 7m) und der entsprechenden mit einem Fluoreszenzfarbstoff markierten Nukleotide für das erfindungsgemäße Verfahren.

Gegenstand der Erfindung ist ferner ein Kit zur Durchführung 20 des Verfahrens das eine Reaktionsoberfläche, zur Durchführung des Verfahrens erforderliche Reaktionslösungen, eine oder mehrere Polymerasen, und Nukleotide (NTs) enthält, von denen ein bis vier mit Fluoreszenzfarbstoffen markiert sind, wobei die NTs ferner strukturell so modifiziert sind (NT' bzw. NTs'), 25 dass die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT' in denselben Strang einzubauen, wobei der Fluoreszenzfarbstoff abspaltbar ist und die strukturelle Modifikation ein abspaltbarer sterisch anspruchsvoller Ligand ist. Bei den 30 Nukleotiden handelt es sich vorzugsweise um die oben genannten erfindungsgemäßen Nukleotide.

Gemäß einer besonderen Ausführungsform der Erfindung enthält das Kit ferner zur Erzeugung von Einzelsträngen aus 35 Doppelsträngen erforderliche Reagenzien, einzelsträngige Nukleinsäuremoleküle, die als PBS in die NSKFs eingeführt werden, Oligonukleotid-Primer, zur Abspaltung der Fluoreszenzfarbstoffe und sterisch anspruchsvollen Liganden

erforderliche Reagenzien und/oder Waschlösungen.

Die erfindungsgemäße Methode dient zur Ermittlung der  
5 Nukleinsäuresequenzen und kann in verschiedenen Bereichen der  
Genetik eingesetzt werden. Dazu zählen insbesondere die  
Bestimmung unbekannter, langer Sequenzen, Analysen von Sequenz-  
Polymorphismen und Punktmutationen sowie die parallele Analyse  
einer großen Zahl an Gensequenzen.

10

Die Vorbereitung des zu analysierenden Materials (einzel- und  
doppelsträngige Nukleinsäuresequenzen) hängt von der  
Aufgabenstellung ab und hat das Ziel, aus einer langen Nuklein-  
säurekette eine Population an relativ kleinen, einzelsträngigen  
15 Nukleinsäurekettenfragmenten (NSKFs) zu bilden, diese Fragmente  
mit einem für den Start der Sequenzierungsreaktion geeigneten  
Primer zu versehen (NSKF-Primer-Komplexe) und auf einer planen  
Oberfläche zu fixieren.

20 Dabei werden einzelne NSKFs auf einer planen Oberfläche in  
einer solchen Weise fixiert, dass eine enzymatische Reaktion an  
diesen Molekülen ablaufen kann. Prinzipiell sind verschiedene  
Arten der Immobilisation möglich, die von der Zielsetzung, der  
Art der NSK und der für die Reaktion eingesetzten Polymerase  
25 abhängen. Die NSKFs werden bei der Immobilisierung bzw. Bindung  
zufällig auf der Oberfläche verteilt, d.h. es muß also nicht  
auf eine exakte Positionierung der einzelnen Ketten geachtet  
werden. NSKF-Primer-Komplexe können über die NSKFs oder Primer  
an die Oberfläche gebunden werden. Die NSKF-Primer-Komplexe  
30 müssen dabei in einer solchen Dichte auf der Oberfläche fixiert  
werden, dass eine eindeutige Zuordnung der später detektierten  
Signale von den eingebauten NT\*s zu einzelnen NSKFs  
gewährleistet ist.

35 Nach der Vorbereitung der NSKFs startet man mit allen auf der  
Oberfläche immobilisierten NSKF-Primer-Komplex-Molekülen die  
Sequenzierungsreaktion. Als Grundlage der Sequenzierung dient  
die Synthese des komplementären Stranges zu jedem einzelnen

- 13 -

gebundenen NSKF. Dabei werden in den neu synthetisierten Strang markierte NTs' eingebaut. Die Polymerase baut nur ein einziges markiertes NT' in die wachsende Kette ein. Dies wird durch die reversible Ankopplung einer zur Termination führenden, sterisch anspruchsvollen Gruppe an die NTs' erreicht. Der Einbau eines weiteren markierten NT' wird dadurch unmöglich gemacht. Diese sterisch anspruchsvolle Gruppe ist vorzugsweise ein Fluoreszenzfarbstoff.

10 Die Sequenzierungsreaktion verläuft in mehreren Zyklen. Ein Zyklus umfasst folgende Schritte:

- a) Zugabe einer Lösung mit markierten Nukleotiden (NTs') und Polymerase zu den gebundenen NSKF-Primer-Komplexen,
- 15 b) Inkubation der gebundenen NSKF-Primer-Komplexe mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
- c) Waschen,
- d) Detektion der Signale von einzelnen Molekülen,
- 20 e) Entfernung der Markierung von den eingebauten Nukleotiden,
- f) Waschen.

Gegebenenfalls erfolgt eine mehrfache Wiederholung des Zyklus (a-f).

25

Die Reaktionsbedingungen des Schrittes (b) in einem Zyklus werden so gewählt, dass die Polymerasen an mehr als 50% der an der Sequenzierungsreaktion beteiligten NSKFs (extensionsfähige NSKF-Primer-Komplexe) in einem Zyklus ein markiertes NT' einbauen können, vorzugsweise an mehr als 90%.

35

Die Anzahl der durchzuführenden Zyklen hängt dabei von der jeweiligen Aufgabenstellung ab, ist theoretisch nicht beschränkt und liegt vorzugsweise zwischen 20 und 5000.

Danach wird für jedes fixierte NSKF seine spezifische Sequenz aus der Reihenfolge der eingebauten NTs' ermittelt.

- 14 -

Aus den überlappenden NSKF-Sequenzen kann in einer Ausführungsform die ursprüngliche NSK-Sequenz rekonstruiert werden ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Dabei sucht man in der gesamten Population von NSKF-Sequenzen nach Übereinstimmungen/Überlappungen in den Sequenzen von NSKFs. Durch diese Übereinstimmungen/Überlappungen kann man die  
10 NSKF in eine Reihe bringen, z.B.:

ACTGTGCGTCCGTATGATGGTCATTCCATG

CATTCCATGGTACGTTAGCTCCTAG

TCCTAGTAAATCGTACC.

15

In der Praxis hat sich bei einer Sequenzierung von unbekannten Sequenzen bewährt, eine Länge der sequenzierten Stücke von mehr als 300 bp zu erreichen. Das erlaubt die Sequenzierung von Genomen aus Eukaryonten im Schrotschuss-Verfahren.

20

Dabei können die Fehler der Methode mit verschiedenen Mitteln erfasst und korrigiert werden. Sämtliche Schritte des Verfahrens können weitgehend automatisiert werden.

25 Durch die Arbeit mit einzelnen Molekülen ergeben sich mehrere Vorteile gegenüber der früher beschriebenen BASS-Methode:

1. Da die Moleküle einzeln detektiert werden, besteht keine Gefahr, dass das Signal durch die Desynchronisation in der  
30 Population fehlerhaft wird. Für jedes fixierte NSKF wird eine eigene Sequenz erstellt. Daher spielt es keine Rolle, ob an einem benachbarten Molekül die Synthese bereits weiter fortgeschritten oder zurückgeblieben ist.
2. Es ist nicht notwendig, Moleküle in einer definierten  
35 Anordnung auf der Oberfläche zu fixieren, da das Signal von einzelnen Molekülen ausgeht und nicht von einer räumlich definierten Population (was bei BASS-Methoden notwendig ist).



#### 4. Detaillierte Beschreibung

##### 4.1 Allgemeine Prinzipien der Reaktion

5

Im folgenden sollen anhand der Sequenzierung eines mehrere Mb langen DNA-Stückes beispielhaft die allgemeinen Prinzipien der Reaktion dargestellt werden (Fig. 1). Der Sequenzierung und der Rekonstruktion von Nukleinsäuresequenzen liegt das Shotgun-  
10 Prinzip zugrunde ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Die Sequenz eines langen DNA-Stücks wird dabei durch  
15 die Sequenzierung kleiner DNA-Fragmente und nachfolgender Rekonstruktion ermittelt. Das zu analysierende Material (1) wird für die Sequenzierungsreaktion vorbereitet, indem es in Fragmente von vorzugsweise 50 bis 1000 bp Länge zerlegt wird (2). Jedes Fragment wird anschließend mit einer  
20 Primerbindungsstelle und einem Primer versehen (3). Dieses Gemisch aus verschiedenen DNA-Fragmenten wird nun auf einer planen Oberfläche fixiert (4). Die nicht gebundenen DNA-Fragmente werden durch einen Waschschrift entfernt. Danach wird die Sequenzierungsreaktion an der gesamten Reaktionsoberfläche  
25 durchgeführt. Diese Reaktion verläuft zyklisch. Im 1. Schritt des Zyklus wird ein mit einem Fluoreszenzfarbstoff markiertes NT' in den wachsenden Strang eingebaut: Dabei wird die Reaktion so gesteuert, dass in jedem Zyklus jeweils nur ein markiertes NT' von einer Polymerase in den wachsenden Strang eingebaut  
30 werden kann. Das wird durch die Verwendung von NTs' erreicht, die eine reversibel gekoppelte, zur Termination führende Gruppe tragen. Der Einbau eines weiteren markierten NT' wird dadurch unmöglich gemacht. Die Polymerase und die markierten NTs' werden gleichzeitig in die Reaktion eingesetzt (5). Danach wird das  
35 Reaktionsgemisch entfernt und die Oberfläche in geeigneter Art und Weise gewaschen (6). Nun folgt ein Detektionsschritt (7): Die Oberfläche wird mit einer für die Einzelmoleküldetektion geeigneten Vorrichtung (bestehend aus Lichtquelle, Mikroskop,

- 16 -

Kamera, Scantisch, Computer mit Steuerungs- und Bilderkennungs- bzw. Bildverarbeitungssoftware) abgescannt und die Signale der einzelnen, eingebauten markierten NTs' identifiziert. Nach dem Detektionsschritt wird die Markierung und die zur Termination  
5 führende Gruppe von allen eingebauten NTs' entfernt (8). Nach einem sich anschließenden Waschschrift kann ein neuer Zyklus beginnen. Zur Rekonstruktion einer größeren ursprünglichen DNA-Sequenz (z.B. mehrere Mb langes DNA-Stück) sollen die DNA-Fragmente einige Hundert NT lang sein, falls man die  
10 Rekonstruktion nach dem Shotgun-Prinzip durchführt ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Da pro Zyklus nur  
15 jeweils ein markiertes NT\* eingebaut wird, sind mindestens 300 Zyklen zur Sequenzierung notwendig.

#### 4.2 Auswahl des Materials

20 Mit Hilfe der erfindungsgemäßen Methode ist es möglich, sowohl vorselektionierte DNA-Sequenzen (z.B. in YAC-, PAC-, oder BAC-Vektoren (R. Anand et al. NAR 1989 v.17 S.3425, H. Shizuya et al. PNAS 1992 v.89 S.8794, "Construction of bacterial  
artificial chromosome libraries using the modified PAC system"  
25 in "Current Protocols in Human genetics" 1996 John Wiley & Sons Inc.) klonierte Abschnitte eines Genoms) als auch nicht vorselektionierte DNA (z.B. genomische DNA, cDNA-Gemische) zu analysieren. Durch eine Vorselektion ist es möglich, im Vorfeld relevante Informationen, wie z.B. Sequenz-Abschnitte aus einem  
30 Genom oder Populationen an Genprodukten, aus der große Menge genetischer Informationen herauszufiltern und damit die Menge der zu analysierenden Sequenzen einzuschränken.

#### 4.3 Vorbereitung des Materials

35

Ziel der Materialvorbereitung ist es, gebundene einzelsträngige NSKFs mit einer Länge von vorzugsweise 50-1000 NTs, einer einzelnen Primerbindungsstelle und einem hybridisierten Primer



- 17 -

(gebundene NSKF-Primer-Komplexe) zu erhalten. Diese NSKF-Primer-Komplexe haben beispielsweise die in Fig. 2 dargestellte Struktur. Im einzelnen können sehr variable Konstruktionen aus dieser allgemeinen Struktur abgeleitet werden. Zur Verbesserung  
5 der Anschaulichkeit folgen nun einige Beispiele, wobei die angeführten Methoden einzeln oder in Kombination eingesetzt werden können.

#### 4.3.1 Erzeugung kurzer Nukleinsäurekettenfragmente (50-1000 10 NTs) (Fragmentierungsschritt)

Wichtig ist, dass die Fragmentierung der NSKs so erfolgt, dass Fragmente erhalten werden, die überlappende Teilsequenzen der Gesamtsequenzen darstellen. Dies wird durch Verfahren erreicht,  
15 bei denen unterschiedlich lange Fragmente als Spaltprodukte in zufällsmäßiger Verteilung entstehen.

Erfindungsgemäß kann die Erzeugung der Nukleinsäurekettenfragmente (NSKFs) durch mehrere Methoden erfolgen, z.B. durch  
20 die Fragmentierung des Ausgangsmaterials mit Ultraschall oder durch Endonukleasen ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press), wie z.B. durch unspezifische Endonukleasegemische. Erfindungsgemäß wird die Ultraschall-Fragmentierung bevorzugt. Man kann die Bedingungen  
25 so einstellen, dass Fragmente mit einer durchschnittlichen Länge von 100 bp bis 1 kb entstehen. Diese Fragmente werden anschließend an ihren Enden durch das Klenow-Fragment (E.coli-Polymerase I) oder durch die T4-DNA-Polymerase aufgefüllt ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor  
30 Laboratory Press).

Ausserdem können aus einer langen NSK unter Verwendung randomisierter Primer komplementäre kurze NSKFs synthetisiert werden. Besonders bevorzugt wird diese Methode bei der Analyse der Gen-  
35 Sequenzen. Dabei werden an der mRNA einzelsträngige DNA-Fragmente mit randomisierten Primern und einer reversen Transkriptase gebildet (Zhang-J et al. Biochem.J. 1999 v.337 S.231, Ledbetter et al. J.Biol.Chem. 1994 v.269 S.31544, Kolls

- 18 -

et al. Anal.Biochem. 1993 v.208 S.264, Decraene et al. Biotechniques 1999 v.27 S.962).

#### 4.3.2 Einführung einer Primerbindungsstelle in das NSKF.

5

Die Primerbindungsstelle (PBS) ist ein Sequenzabschnitt, der eine selektive Bindung des Primers an das NSKF ermöglichen soll.

10 In einer Ausführungsform können die Primerbindungsstellen unterschiedlich sein, so dass mehrere unterschiedliche Primer verwendet werden müssen. In diesem Fall können bestimmte Sequenzabschnitte der Gesamtsequenz als natürliche PBSs für spezifische Primer dienen. Diese Ausführungsform ist besonders  
15 für die Untersuchung bereits bekannter SNP-Stellen geeignet, s. Beispiel 5 "SNP-Analyse mit sequenzspezifischen Primern".

In einer anderen Ausführungsform ist es aus Gründen der Vereinfachung der Analyse günstig, wenn eine einheitliche  
20 Primerbindungsstelle in allen NSKFs vorhanden ist. Gemäß einer bevorzugten Ausführungsform der Erfindung werden die Primerbindungsstellen daher in die NSKFs extra eingeführt. Auf diese Weise können Primer mit einheitlicher Struktur für die Reaktion eingesetzt werden.

25

Im folgenden wird diese Ausführungsform detailliert beschrieben.

Die Zusammensetzung der Primerbindungsstelle ist nicht einge-  
30 schränkt. Ihre Länge beträgt vorzugsweise zwischen 20 und 50 NTs. Die Primerbindungsstelle kann eine funktionelle Gruppe zur Immobilisation des NSKF tragen. Diese funktionelle Gruppe kann z.B. eine Biotingruppe sein.

35 Als Beispiel für die Einführung einer einheitlichen Primerbindungsstelle werden im folgenden die Ligation und das Nukleotid-Tailing an DNA-Fragmente beschrieben.

- 19 -

## a) Ligation:

Dabei wird ein doppelsträngiger Oligonukleotidkomplex mit einer Primerbindungsstelle verwendet (Fig. 3a). Dieser wird mit kommerziell erhältlichen Ligasen an die DNA-Fragmente ligiert 5 ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press). Es ist wichtig, dass nur eine einzige Primerbindungsstelle an das DNA-Fragment ligiert wird. Das erreicht man z.B. durch eine Modifikation einer Seite des Oligonukleotidkomplexes an beiden Strängen (Fig. 3b). Die 10 Resultate nach der Ligation bzw. nach anschließender Denaturierung sind in Fig. 3c und 3d dargestellt. Die modifizierenden Gruppen am Oligonukleotidkomplex können zur Immobilisation dienen. Die Synthese und die Modifikation eines solchen Oligonukleotidkomplexes kann nach standardisierten 15 Vorschriften durchgeführt werden. Zur Synthese kann z.B. der DNA-Synthesizer 380 A Applied Biosystems verwendet werden. Oligonucleotide mit einer bestimmten Zusammensetzung mit oder ohne Modifikationen sind aber auch als Auftragssynthese kommerziell erhältlich, z.B. von MWG-Biotech GmbH, Germany.

20

## b) Nukleotid-Tailing:

Statt der Ligation mit einem Oligonukleotid kann man mit einer terminalen Deoxynucleotidyltransferase mehrere (z.B. zwischen 10 und 20) Nukleosid-monophosphate an das 3'-Ende eines ss-DNA- 25 Fragments anknüpfen ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press, "Method in Enzymology" 1999 v.303, S.37-38) (Fig. 4), z.B. mehrere Guanosin-Monophosphate ((G)<sub>n</sub>-Tailing genannt). Das entstehende Fragment wird zur Bindung des Primers, in diesem Beispiel eines (C)<sub>n</sub>- 30 Primers, verwendet.

## 4.3.3 Einzelstrang-Vorbereitung

Für die Sequenzierungsreaktion werden einzelsträngige NSKFs 35 benötigt. Falls das Ausgangsmaterial in doppelsträngiger Form vorliegt, gibt es mehrere Möglichkeiten, aus doppelsträngiger DNA eine einzelsträngige Form zu erzeugen (z.B. Hitze-Denaturierung oder Alkali-Denaturierung) ("Molecular cloning"

- 20 -

1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press).

#### 4.3.4 Primer für die Sequenzierungsreaktion

5 Dieser hat die Funktion, den Start an einer einzigen Stelle des  
NSKF zu ermöglichen. Er bindet an die Primerbindungsstelle im  
NSKF. Die Zusammensetzung und die Länge des Primers sind nicht  
eingeschränkt. Außer der Startfunktion kann der Primer auch  
andere Funktionen übernehmen, wie z.B. eine Verbindung zur  
10 Reaktionsoberfläche zu schaffen. Primer sollten so an die Länge  
und Zusammensetzung der Primerbindungsstelle angepaßt werden,  
dass der Primer den Start der Sequenzierungsreaktion mit der  
jeweiligen Polymerase ermöglicht.

15 Bei der Verwendung unterschiedlicher, beispielsweise natürlich  
in der ursprünglichen Gesamtsequenz vorkommender  
Primerbindungsstellen, werden die für die jeweilige  
Primerbindungsstelle sequenzspezifischen Primer verwendet. In  
diesem Fall wird für die Sequenzierung ein Primergemisch  
20 eingesetzt.

Bei einer einheitlichen, beispielsweise durch die Ligation an  
die NSKFs angekoppelten Primerbindungsstelle wird ein  
einheitlicher Primer verwendet.

25

Vorzugsweise beträgt die Länge des Primers zwischen 6 und 100  
NTs, optimalerweise zwischen 15 und 30 NTs. Der Primer kann  
eine Funktionsgruppe tragen, die zur Immobilisierung des NSKF  
dient, beispielsweise ist eine solche Funktionsgruppe eine  
30 Biotingruppe (s. Abschnitt Immobilisierung). Sie soll die  
Sequenzierung nicht stören. Die Synthese eines solchen Primers  
kann z.B. mit dem DNA-Synthesizer 380 A Applied Biosystems  
ausgeführt werden oder aber als Auftragssynthese bei einem  
kommerziellen Anbieter, z.B. MWG-Biotech GmbH, Germany erstellt  
35 werden).

Der Primer kann vor der Hybridisierung an die zu analysierenden  
NSKFs auf der Oberfläche mit verschiedenen Techniken fixiert

- 21 -

oder direkt auf der Oberfläche synthetisiert werden, beispielsweise nach (McGall et al. US Patent 5412087, Barrett et al. US Patent 5482867, Mirzabekov et al. US Patent 5981734, "Microarray biochip technology" 2000 M.Schena Eaton Publishing, 5 "DNA Microarrays" 1999 M. Schena Oxford University Press, Fodor et al. Science 1991 v.285 S.767, Timofeev et al. Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353, Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR 1992 v.20 S.1679).

10

Die Primer werden auf der Oberfläche beispielsweise in einer Dichte zwischen 10 bis 100 pro 100  $\mu\text{m}^2$ , 100 bis 10.000 pro 100  $\mu\text{m}^2$  oder 10.000 bis 1.000.000 pro 100  $\mu\text{m}^2$  gebunden.

15 Der Primer oder das Primergemisch wird mit NSKFs unter Hybridisierungsbedingungen inkubiert, die ihn selektiv an die Primerbindungsstelle des NSKF binden lassen. Diese Primer-Hybridisierung (Annealing) kann vor (1), während (2) oder nach (3) der Bindung der NSKFs an die Oberfläche erfolgen. Die 20 Optimierung der Hybridisierungsbedingungen hängt von der genauen Struktur der Primerbindungsstelle und des Primers ab und läßt sich nach Rychlik et al. NAR 1990 v.18 S.6409 berechnen. Im folgenden werden diese Hybridisierungsbedingungen als standardisierte Hybridisierungsbedingungen bezeichnet.

25

Falls eine für alle NSKFs gemeinsame Primerbindungsstelle mit bekannter Struktur beispielsweise durch Ligation eingeführt wird, können Primer mit einheitlicher Struktur eingesetzt werden. Die Primerbindungsstelle kann an ihrem 3'-Ende eine 30 funktionelle Gruppe tragen, die z.B. zur Immobilisation dient. Beispielsweise ist diese Gruppe eine Biotin-Gruppe. Der Primer hat eine zur Primerbindungsstelle komplementäre Struktur.

Ein Beispiel einer Primerbindungsstelle und eines Primers ist 35 nachfolgend dargestellt.

5' TAATACGACTCACTATAGG3'	Primer (T7-19-Primer)
Biotin-3' ATTATGCTGAGTGATATCC5'	Primerbindungsstelle



#### 4.3.5 Fixierung von NSKF-Primer-Komplexe an die Oberfläche (Bindung bzw. Immobilisierung von NSKFs).

5 Ziel der Fixierung (Immobilisierung) ist es, NSKF-Primer-Komplexe auf einer geeigneten planen Oberfläche in einer Art und Weise zu fixieren, dass eine zyklische enzymatische Sequenzierungsreaktion ablaufen kann. Dies kann beispielsweise durch Bindung des Primers (s.o.) oder des NSKF an die  
10 Oberfläche erfolgen.

Die Reihenfolge der Schritte bei der Fixierung von NSKF-Primer-Komplexen kann variabel sein:

- 1) Die NSKF-Primer-Komplexe können zunächst in einer Lösung  
15 durch Hybridisierung (Annealing) gebildet und anschließend an die Oberfläche gebunden werden.
  - 2) Primer können zunächst auf einer Oberfläche gebunden werden und NSKFs anschließend an die gebundenen Primer hybridisiert werden, wobei NSKF-Primer-Komplexe entstehen  
20 (NSKFs indirekt an die Oberfläche gebunden)
  - 3) Die NSKFs können zunächst an die Oberfläche gebunden werden (NSKFs direkt an die Oberfläche gebunden) und im anschließenden Schritt die Primer an die gebundenen NSKFs hybridisiert werden, wobei NSKF-Primer-Komplexe entstehen.
- 25 Die Immobilisierung der NSKFs an die Oberfläche kann daher durch direkte oder indirekte Bindung erfolgen.

Oberfläche und Reaktionsoberfläche sind vorliegend als gleichwertige Begriffe aufzufassen, außer wenn explizit auf  
30 eine andere Bedeutung hingewiesen wird. Als Reaktionsoberfläche dient die Oberfläche einer festen Phase eines beliebigen Materials. Dieses Material ist vorzugsweise enzymatischen Reaktionen gegenüber inert und verursacht keine Störungen der Detektion. Silicon, Glas, Keramik, Kunststoff (z.B.  
35 Polycarbonate oder Polystyrole), Metall (Gold, Silber, oder Aluminium) oder beliebiges anderes Material, das diesen funktionellen Anforderungen genügt, kann verwendet werden. Vorzugsweise ist die Oberfläche nicht verformbar, denn sonst

ist mit einer Verzerrung der Signale bei der wiederholten Detektion zu rechnen.

Falls eine gelartige feste Phase (Oberfläche eines Gels) verwendet wird, so kann dieses Gel z.B. ein Agarose- oder Polyacrylamidgel sein. Das Gel ist vorzugsweise für Moleküle mit einer Molekularmasse unter 5000 Da frei passierbar (beispielsweise kann ein 1 bis 2% Agarose-Gel oder 10 bis 15% Polyacrylamid Gel verwendet werden). Eine solche Geloberfläche hat anderen festen Oberflächen gegenüber den Vorteil, dass es zu einer wesentlich geringeren unspezifischen Bindung von NT\*s an die Oberfläche kommt. Durch die Bindung der NSKF-Primer-Komplexe auf der Oberfläche ist die Detektion der Fluoreszenzsignale von eingebauten NTs' möglich. Die Signale von freien NTs' werden nicht detektiert, weil sie nicht an das Material des Gels binden und somit nicht immobilisiert werden. Das Gel ist vorzugsweise auf einer festen Unterlage befestigt (Fig. 5a). Diese feste Unterlage kann Silicon, Glas, Keramik, Kunststoff (z.B. Polycarbonate oder Polystyrole), Metall (Gold, Silber, oder Aluminium) oder beliebiges anderes Material sein. Die Dicke des Gels beträgt vorzugsweise nicht mehr als 0,1 mm. Die Geldicke ist vorzugsweise größer als die einfache Tiefenschärfe des Objektivs sein, damit unspezifisch an die feste Unterlage gebundene NTs' nicht in die Fokusebene gelangen und damit detektiert werden. Wenn die Tiefenschärfe z.B. 0,3 µm beträgt, so liegt die Geldicke vorzugsweise zwischen 1 µm und 100 µm. Die Oberfläche kann als eine kontinuierliche Oberfläche oder als diskontinuierliche, aus einzelnen kleinen Bestandteilen (z.B. Agarose-Kügelchen) zusammengesetzte Oberfläche hergestellt werden (Fig.5b). Die Reaktionsoberfläche muß groß genug sein, um die notwendige Anzahl der NSKFs bei entsprechender Dichte immobilisieren zu können. Die Reaktionsoberfläche sollte vorzugsweise nicht größer als 20 cm<sup>2</sup> sein.

Die verschiedenen Zyklusschritte erfordern einen Austausch der unterschiedlichen Reaktionslösungen über der Oberfläche. Die Reaktionsoberfläche ist vorzugsweise Bestandteil eines Reaktionsgefäßes. Das Reaktionsgefäß ist wiederum vorzugsweise



- 24 -

Bestandteil einer Reaktionsapparatur mit Durchflußvorrichtung. Die Durchflußvorrichtung ermöglicht einen Austausch der Lösungen im Reaktionsgefäß. Der Austausch kann mit einer durch einen Computer gesteuerten Pumpvorrichtung oder manuell  
5 erfolgen. Wichtig dabei ist, dass die Oberfläche nicht austrocknet. Vorzugsweise beträgt das Volumen des Reaktionsgefäßes weniger als 50 µl. Idealerweise beträgt sein Volumen weniger als 1 µl. Ein Beispiel eines solchen Durchflußsystems ist in Fig.6 gegeben.

10

Falls die Fixierung der NSKF-Primer-Komplexe auf der Oberfläche über die NSKFs erfolgt, kann dies beispielsweise durch die Bindung der NSKFs an einem der beiden Ketten-Enden erfolgen. Dies kann durch entsprechende kovalente, affine oder andere  
15 Bindungen erreicht werden. Es sind viele Beispiele der Immobilisierung von Nukleinsäuren bekannt (McGall et al. US Patent 5412087, Nikiforov et al. US Patent 5610287, Barrett et al. US Patent 5482867, Mirzabekov et al. US Patent 5981734, "Microarray biochip technology" 2000 M.Schena Eaton Publishing,  
20 "DNA Microarrays" 1999 M. Schena Oxford University Press, Rasmussen et al. Analytical Biochemistry v.198, S.138, Allemand et al. Biophysical Journal 1997, v.73, S.2064, Trabesinger et al. Analytical Chemistry 1999, v.71, S.279, Osborne et al. Analytical Chemistry 2000, v.72, S.3678, Timofeev et al.  
25 Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353, Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR 1992 v.20 S.1679). Die Fixierung kann auch durch eine unspezifische Bindung, wie z.B. durch Austrocknung der NSKFs enthaltenden Probe auf der planen Oberfläche erreicht  
30 werden.

Die NSKFs werden auf der Oberfläche beispielsweise in einer Dichte zwischen 10 und 100 NSKFs pro 100 µm<sup>2</sup>, 100 bis 10.000 pro 100 µm<sup>2</sup>, 10.000 bis 1.000.000 pro 100µm<sup>2</sup> gebunden.

35

Die für die Detektion notwendige Dichte von extensionsfähigen NSKF-Primer-Komplexen beträgt ca. 10 bis 100 pro 100 µm<sup>2</sup>. Sie kann vor, während oder nach der Hybridisierung der Primer an

die Genprodukte erreicht werden.

Beispielhaft werden im folgenden einige Methoden zur Bindung von NSKF-Primer-Komplexen näher dargestellt: In einer Ausführungsform erfolgt die Immobilisierung der NSKFs über Biotin-Avidin oder Biotin-Streptavidin-Bindung. Dabei wird Avidin oder Streptavidin auf der Oberfläche kovalent gebunden, das 5'-Ende des Primers enthält Biotin. Nach der Hybridisierung der markierten Primer mit den NSKFs (in Lösung) werden diese auf der mit Avidin/Streptavidin beschichteten Oberfläche fixiert. Die Konzentration der mit Biotin markierten Hybridisierungs-Produkte sowie die Zeit der Inkubation dieser Lösung mit der Oberfläche wird so gewählt, dass eine für die Sequenzierung geeignete Dichte bereits in diesem Schritt erreicht wird.

15

In einer anderen bevorzugten Ausführungsform werden die für die Sequenzierungsreaktion geeigneten Primer vor der Sequenzierungsreaktion auf der Oberfläche mit geeigneten Methoden fixiert (s.o.). Die einzelsträngigen NSKFs mit jeweils einer Primerbindungsstelle pro NSKF werden damit unter Hybridisierungsbedingungen inkubiert (Annealing). Dabei binden sie an die fixierten Primer und werden dadurch gebunden (indirekte Bindung), wobei Primer-NSKF-Komplexe entstehen. Die Konzentration der einzelsträngigen NSKFs und die Hybridisierungsbedingungen werden so gewählt, dass man eine für die Sequenzierung geeignete Immobilisationsdichte von 10 bis 100 extensionsfähigen NSKF-Primer-Komplexen pro  $100 \mu\text{m}^2$  erreicht. Nach der Hybridisierung werden ungebundene NSKFs durch einen Waschschrift entfernt. Bei dieser Ausführungsform wird eine Oberfläche mit einer hohen Primerdichte bevorzugt, z.B. ca. 1.000.000 Primer pro  $100 \mu\text{m}^2$  oder noch höher, da die gewünschte Dichte an NSKF-Primer-Komplexen schneller erreicht wird, wobei die NSKFs nur an einen Teil der Primer binden.

35 In einer anderen Ausführungsform werden die NSKFs an die Oberfläche direkt gebunden (s.o.) und anschließend mit Primern unter Hybridisierungsbedingungen inkubiert. Bei einer Dichte von ca. 10 bis 100 NSKFs pro  $100 \mu\text{m}^2$  wird man versuchen alle

- 26 -

verfügbaren NSKFs mit einem Primer zu versehen und für die Sequenzierungsreaktion verfügbar zu machen. Dies kann z.B. durch hohe Primerkonzentration, beispielsweise 1 bis 100 mmol/l, erreicht werden. Bei einer höheren Dichte der fixierten 5 NSKFs auf der Oberfläche, beispielsweise 10.000 bis 1.000.000 pro 100µm<sup>2</sup>, kann die für die optische Detektion notwendige Dichte der NSKF-Primer-Komplexe während der Primer-Hybridisierung erreicht werden. Dabei sind die Hybridisierungsbedingungen (z.B. Temperatur, Zeit, Puffer, 10 Primerkonzentration) so zu wählen, dass die Primer nur an einen Teil der immobilisierten NSKFs binden, s. Beispiel 5,6.

Falls die Oberfläche einer festen Phase (z.B. Silikon oder Glas) zur Immobilisation verwendet wird, wird vorzugsweise eine 15 Blockierungslösung auf die Oberfläche vor dem Schritt (a) in jedem Zyklus gebracht, die zur Vermeidung einer unspezifischen Adsorption von NTs' an der Oberfläche dient. Diese Bedingungen für eine Blockierlösung erfüllt beispielsweise eine Albuminlösung (BSA) mit einem pH-Wert zwischen 8 und 10.

20

#### 4.4 Wahl der Polymerase

Als Polymerasen eignen sich prinzipiell alle DNA-abhängigen DNA-Polymerasen ohne 3'-5' Exonuklease-Aktivität (DNA- 25 Replication" 1992 Ed. A.Kornberg, Freeman and company NY), z.B. modifizierte T7-Polymerase vom Typ "Sequenase Version 2" (Amersham Pharmacia Biotech), 3'-5' exonuklease freies Klenow Fragment der DNA-Polymerase I (Amersham Pharmacia Biotech), Polymerase Beta verschiedenen Ursprungs (Animal Cell DNA 30 Polymerases" 1983, Fry M., CRC Press Inc., kommerziell erhältlich bei Chimerx) thermostabile Polymerasen wie beispielsweise Taq-Polymerase (GibcoBRL), proHATM Polymerase (Eurogentech).

35 Polymerasen mit 3'-5' Exonuklease-Aktivität können eingesetzt werden (z.B. Klenow-Fragment der E.coli-Polymerase I), sofern Reaktionsbedingungen gewählt werden, die vorhandene 3'-5' Exonuklease-Aktivität unterdrücken, wie z.B. ein niedriger pH-

- 27 -

Wert (pH 6.5) beim Klenow-Fragment (Lehman and Richardson, J. Biol. Chem. 1964 v.239 S.233) oder Zugabe von NaF zur Einbaureaktion. Eine andere Möglichkeit besteht in der Verwendung von NTs' mit einer Phosphorothioate-Verbindung 5 (Kunkel et al. PNAS 1981, v.78 S.6734). Dabei werden eingebaute NTs' von der 3'-5' Exonuklease-Aktivität der Polymerase nicht angegriffen. Im folgenden werden all diese Polymerasearten als "Polymerase" bezeichnet.

## 10 4.5 Chemie

### 4.5.1 Allgemeines Prinzip

Für die Sequenzierungsreaktion bei hoch paralleler 15 Sequenzanalyse an einzelnen Nukleinsäure-Molekülen (parallele Analyse von bis zu 10.000.000 NSKF-Sequenzen) ist wichtig, dass jedes eingebaute NT' identifiziert wird. Eine Voraussetzung dafür ist, dass nur ein einziges NT' pro Zyklus in die Nukleinsäurekette eingebaut wird. Falls eine Polymerase mehrere 20 NTs' nacheinander im selben Zyklus einbaut, so führt dies zu einem Fehler in der Sequenzermittlung. Aus diesem Grund muß man den Einbau der NTs' steuern.

Beispielsweise wurden in der BASS-Methode reversible 3'-OH modifizierte NTs beschrieben (Dower US Patent 5.547.839, Canard 25 et al. US Patent 5.798.210, Rasolonjatovo Nucleosides & Nucleotides 1999, v.18 S.1021, Metzker et al. NAR 1994, v.22, S.4259, Welch et al. Nucleosides & Nucleotides 1999, v.18, S.197). Die Spaltung soll dabei unter milden Bedingungen photochemisch (Dower US Patent 5.547.839, Welch et al. 30 Nucleosides & Nucleotides 1999, v.18, S.197) oder chemisch (Canard et al. US Patent 5.798.210, Rasolonjatovo Nucleosides & Nucleotides 1999, v.18 S.1021) erfolgen.

Die Synthese der 3'-OH-modifizierten photochemisch spaltbaren 35 NTs' ist sehr aufwendig. Die Polymerasen weisen eine sehr unterschiedliche Affinität zu diesen Nukleotidanalogen auf, so dass die Nukleinsäuresynthese sehr ungleichmäßig bzw. an vielen DNA-Stellen gar nicht abläuft (Metzker et al. NAR 1994 v.22 S.4259,

- 28 -

Welch et al. Nucleosides & Nucleotides 1999, v.18 S.197). Aus diesen Gründen eignen sich diese Analoga nicht oder nur sehr eingeschränkt für die Sequenzierungsreaktion. Eine spaltbare 3'-Ester-Verknüpfung (Canard et al. US Patent 5.798.210) kommt 5 als Grundlage für eine reversible Termination der Synthese auch nicht in Betracht. Die meisten Polymerasen spalten bei Verfügbarkeit eines nächsten komplementären NT 3'-OH-Ester-Verbindungen, so dass die an die 3'-OH-Gruppe gebundene Markierung in die Lösung freigesetzt wird und nicht mehr als 10 Terminator wirken kann (Rasolonjatovo et al. Nucleosides & Nucleotides 1999, v.18 S.1021, Canard et al. PNAS 1995 v.92 S.10859). In Positionen, an denen eine Polymerase mehrere gleiche NTs nacheinander einbauen kann, führt das zu einem fehlerhaften Signal. Im Abschlußbericht des BMBF- 15 Verbundprojekts „Sequenzierung mit Multiplexfarbstoffen und Kapillarelektrophorese“ G. Sagner, 1999, wurde berichtet, dass Modifikationen der 3'-Position von Nukleotiden zur Aufhebung ihrer Substrateigenschaften für Polymerasen geführt haben.

20 Die Schwierigkeiten bei der Entwicklung passender NT-Analoga für das Verfahren basieren auf folgenden Rahmenbedingungen:

- 1) Die Reaktion muss so gesteuert werden, dass die Polymerase NT\*s einzeln einbaut (Stop des weiteren Einbaus).
- 2) NT\*s müssen einen Farbstoff tragen, der den Anforderungen 25 der Detektion genügt.
- 3) Der Farbstoff muß unter milden Bedingungen abspaltbar sein, so dass weder die NSKF-Primer-Komplexe, noch einzelne Komponenten des Systems beschädigt werden.
- 4) Die Abspaltung muss möglichst schnell und quantitativ 30 erfolgen.
- 5) Der Stop des Einbaus muss reversibel sein und unter milden Bedingungen aufgehoben werden können.

Bis jetzt wurde keine praktisch brauchbare Lösung für diese 35 Probleme in der einschlägigen Literatur vorgestellt.

Durch die vorliegende Erfindung werden nunmehr die im Stand der Technik bekannten Probleme erstmals gelöst. Erfindungsgemäß



- 29 -

werden für die Sequenzierung NTs' mit einer sterisch anspruchsvollen Gruppe an der Base verwendet.

Eine an die Base gekoppelte sterisch anspruchsvolle Gruppe kann zur Behinderung der weiteren Synthese führen, wobei diese Behinderung in der Fachliteratur als unerwünschte Eigenschaft modifizierter NTs bei der Markierung von Nukleinsäuren angesehen wird. Biotin, Digoxigenin und Fluoreszenzfarbstoffe wie Fluoreszein, Tetramethylrhodamine, Cy3-Farbstoff stellen Beispiele einer solchen sterisch anspruchsvollen Gruppe dar (Zhu et al. Cytometry 1997, v.28, S.206, Zhu et al. NAR 1994, v.22, S.3418, Gebeyehu et al., NAR 1987, v.15, S.4513, Wiemann et al. Analytical Biochemistry 1996, v.234, S.166, Heer et al. BioTechniques 1994 v.16 S.54).

15

Bei der Sequenzierungsreaktion im erfindungsgemäßen Verfahren werden markierte NTs' mit einer Polymerase und Nukleinsäureketten inkubiert. Die NTs' tragen dabei eine an die Base reversibel gekoppelte sterisch anspruchsvolle Gruppe. Wenn ein Reaktionsgemisch, das nur modifizierte NTs' enthält, in der Reaktion eingesetzt wird, dann kann die Polymerase nur ein einziges NT' einbauen. Der Einbau eines nächsten NT' wird sterisch gehemmt. Diese NTs' treten somit als Terminatoren der Synthese auf. Nach der Entfernung der sterisch anspruchsvollen Gruppe kann das nächste komplementäre NT' eingebaut werden. Weil diese NTs' kein absolutes Hindernis zur weiteren Synthese darstellen, sondern nur für den Einbau eines weiteren markierten NT', werden sie als Semiterminatoren bezeichnet.

Der Unterschied zur 3'-OH Terminatoren-Methode besteht darin, dass nicht eine Blockade der für die Synthese notwendigen 3'-OH Gruppe angestrebt wird, sondern eine an die Base geknüpfte Gruppe als sterisches Hindernis für den weiteren Einbau genutzt wird. Die 3'-OH Gruppe bleibt dabei die ganze Zeit frei.

35

#### 4.5.2 Allgemeine Struktur des NT'

Ihre gemeinsamen Merkmale sind in Fig. 7a,b,d dargestellt.

- 30 -

Diese Struktur ist dadurch charakterisiert, dass an der Base über einen spaltbaren Linker (A-E) eine sterische Gruppe (D) und der Fluoreszenzmarker (F) gebunden sind.

5 Als Grundlage für die NTs' dienen Deoxynukleosid-Triphosphate mit Adenosin (A), Guanosin (G), Cytidin (C) und Thymidin (T) als Nukleosidrest. Anstelle von Thymidin wird bevorzugt Uridin als Nukleosidrest verwendet. Anstelle von Guanosin kann Inosin verwendet werden.

10

#### 4.5.3 Marker, Fluorophore

Jede Base ist mit einem für sie charakteristischen Marker (F) markiert (Fig. 7). Der Marker ist ein fluoreszierender Farbstoff. Mehrere Faktoren beeinflussen die Wahl des Fluoreszenzfarbstoffes. Die Wahl ist nicht eingeschränkt, sofern der Farbstoff folgenden Anforderungen genügt:

a) Die verwendete Detektionsapparatur muß diesen Marker als  
20 einziges Molekül gebunden an DNA unter milden Bedingungen (vorzugsweise Reaktionsbedingungen) identifizieren können. Die Farbstoffe haben vorzugsweise große Photostabilität. Ihre Fluoreszenz wird vorzugsweise von der DNA nicht oder nur unwesentlich gequenchet.

25

b) Der an das NT gebundene Farbstoff darf keine irreversible Störung der enzymatischen Reaktion verursachen.

c) mit dem Farbstoff markierte NTs' müssen von der Polymerase in  
30 die Nukleinsäurekette eingebaut werden.

d) Bei einer Markierung mit verschiedenen Farbstoffen sollen diese Farbstoffe keine beträchtlichen Überlappungen in ihren Emissionsspektren aufweisen.

35

Im Rahmen der vorliegenden Erfindung verwendbare Fluoreszenzfarbstoffe sind in "Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular



Probes mit Strukturformeln zusammengestellt. Erfindungsgemäß werden vorzugsweise folgende Farbstoffklassen als Marker eingesetzt: Cyanin-Farbstoffe und deren Abkömmlinge (z.B. Cy2, Cy3, Cy5, Cy7 Amersham Pharmacia Biotech, Waggoner US-Patent 5 5.268.486), Rhodamine und deren Abkömmlinge (z.B. TAMRA, TRITC, RG6, R110, ROX, Molecular Probes, s. Handbuch), Xanthene-Derivate (z.B. Alexa 568, Alexa 594, Molecular Probes, Mao et al. US-Patent 6.130.101). Diese Farbstoffe sind kommerziell erhältlich.

10

Dabei kann man je nach spektralen Eigenschaften und vorhandener Apparatur entsprechende Farbstoffe auswählen. Die Farbstoffe werden an den Linker z.B. über Thiocyanat- oder Ester-Bindung gekoppelt ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes, Jameson et al. Methods in Enzymology 1997 v.278 S.363, Waggoner Methods in Enzymology 1995 v.246 S.362), s. Beispiele 1 und 2.

#### 4.5.4 Natur der sterisch anspruchsvollen Gruppe

20

Die Gruppe (D) (Fig. 7a,b,d) stellt ein Hindernis für den Einbau eines weiteren komplementären markierten NT durch eine Polymerase dar. Biotin, Digoxigenin und Fluoreszenzfarbstoffe stellen Beispiele einer solchen sterisch anspruchsvollen Gruppe dar (Zhu et al. Cytometry 1997, v.28, S.206, Zhu et al. NAR 1994, v.22, S.3418, Gebeyehu et al., NAR 1987, v.15, S.4513, Wiemann et al. Analytical Biochemistry 1996, v.234, S.166, Heer et al. BioTechniques 1994 v.16 S.54). Die chemische Struktur dieser Gruppe ist nicht eingeschränkt, sofern sie den Einbau 25 des markierten NT, an das sie geknüpft ist, nicht wesentlich stört und keine irreversible Störung der enzymatischen Reaktion verursacht.

Diese Gruppe kann als selbständiger Teil im Linker (7a) auftreten oder mit dem Farbstoff (7b) oder der spaltbaren Gruppe (7d) identisch sein. Durch die Spaltung des Linkers wird diese sterisch anspruchsvolle Gruppe (D) nach der Detektion des Signals entfernt, so dass die Polymerase ein weiteres markiertes 35

- 32 -

NT\* einbauen kann. Bei einer Struktur wie in 7d wird die sterische Gruppe durch die Spaltung beseitigt.

5 In einer bevorzugten Ausführungsform übernimmt der Fluoreszenzfarbstoff die Funktion einer solchen sterisch anspruchsvollen Gruppe, so dass ein markiertes Nukleotid eine in Fig. 7b,k,l dargestellte Struktur aufweist.

10 In einer anderen bevorzugten Ausführungsform übernimmt die photolabile spaltbare Gruppe die Funktion einer solchen sterisch anspruchsvollen Gruppe (Fig. 7d).

#### 4.5.5 Linker

15

Der Marker (Fluoreszenzfarbstoff) ist an die Base vorzugsweise über einen Abstandhalter unterschiedlicher Länge, einen sog. Linker, gebunden. Beispiele für Linker sind in Fig. 7e,f,h,i,j,k,l,m gegeben. Vorzugsweise ist dieser Linker an  
20 einer der Stellen an die Base gebunden, die nicht an der Basenpaarung teilnimmt. Im bevorzugten Fall sind die Stellen, an die der Linker gebunden ist: die 5-Position im Pyrimidinring und die 7-Position oder 8-Position im Purinring. Beispiele der Ankoppelung eines Linkers an die Base können aus folgenden  
25 Quellen entnommen werden (Hobbs et al. US Patent 5.047.519, Khan et al. US Patent 5.821.356, Hanna M. Method in Enzymology 1996 v.274, S.403, Zhu et al. NAR 1994 v.22 S.3418, Herman et al. Methods in Enzymology 1990 v.184 S.584, J.L.Ruth et al. Molecular Pharmacology 1981 v.20 S.415, L. Ötvös et al. NAR  
30 1987 v.15 S.1763, G.E.Wright et al. Pharmac Ther. 1990 v.47, S.447, „Nucleotide Analogs; Synthesis and Biological Function“ K.H. Scheit 1980, Wiley-Interscience Publication, "Nucleic acid chemistry" Ed. L.B.Townsend, v.1-4, Wiley-Interscience Publication, "Chemistry of Nucleosides and Nucleotides" Ed.  
35 L.B.Townsend, v.1-3, Plenum Press).

Die gesamte Länge des Linkers kann variieren. Sie entspricht der Anzahl der Kohlenstoff-Atome in den Abschnitten A, C, E

(Fig. 7a,b,d) und liegt vorzugsweise zwischen 3 und 20. Optimalerweise beträgt sie zwischen 4 und 10 Atomen. Die chemische Zusammensetzung des Linkers (Abschnitte A,C,E in Fig. 7a,b,d) ist nicht eingeschränkt, sofern sie unter 5 Reaktionsbedingungen stabil bleibt und keine Störung der enzymatischen Reaktion verursacht.

#### 4.5.6 Spaltbare Verbindung, Spaltung

10 Der Linker trägt eine spaltbare Verbindung oder spaltbare Gruppe (Abschnitt (B) in Fig. 7a,b,d und Abschnitt (C) in Fig. 7k,l). Diese spaltbare Verbindung ermöglicht die Entfernung des Markers und des sterischen Hindernisses am Ende jedes Zyklus. Ihre Wahl ist nicht eingeschränkt, sofern sie unter den Bedin-  
15 gungen der enzymatischen Sequenzierungsreaktion stabil bleibt, keine irreversible Störung der Polymerase verursacht und unter milden Bedingungen abgespalten werden kann. Unter "milden Bedingungen" sind solche Bedingungen zu verstehen, die den NSKF-Primer-Komplex nicht zerstören, wobei z.B. der pH-Wert  
20 vorzugsweise zwischen 3 und 11 liegt, die Temperatur zwischen 0°C und einem Temperaturwert (x). Dieser Temperaturwert (x) hängt von der T<sub>m</sub> des NSKF-Primer-Komplexes (T<sub>m</sub> ist "melting Point") und wird beispielsweise als T<sub>m</sub> (NSKF-Primer-Komplex) minus 5°C errechnet (z.B. T<sub>m</sub> ist 47°C, dann liegt die maximale  
25 Temperatur bei 42°C; unter diesen Bedingungen eignen sich besonders Ester-, Thioester-, Disulfid-Verbindungen und photolabile Verbindungen als spaltbare Verbindungen).

Vorzugsweise gehört die genannte Verbindung zu chemisch oder  
30 enzymatisch spaltbaren oder photolabilen Verbindungen. Als Beispiele von chemisch spaltbaren Gruppen sind Ester-, Thioester- und Disulfid-Verbindungen bevorzugt (Fig. 7k,l) („Chemistry of protein conjugation and crosslinking“ Shan S. Wong 1993 CRC Press Inc., Herman et al. Method in Enzymology  
35 1990 v.184 S.584, Lomant et al. J.Mol.Biol. 1976 v.104 243, "Chemistry of carboxylic acid and esters" S.Patai 1969 Interscience Publ.). Beispiele für photolabile Verbindungen

- 34 -

(Fig. 7m) können in folgenden Literaturstellen gefunden werden:  
"Protective groups in organic synthesis" 1991 John Wiley & Sons, Inc., V. Pillai Synthesis 1980 S.1, V. Pillai Org. Photochem. 1987 v.9 S.225, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese" H.Giegrich, 1996, Konstanz, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese" S.M.Bühler, 1999, Konstanz) ..

- 10 Die Position der spaltbaren Verbindung/Gruppe im Linker ist vorzugsweise nicht weiter als 10 Atome von der Base entfernt, noch bevorzugter nicht weiter als 3 Atome. Besonders bevorzugt liegt die spaltbare Verbindung oder Gruppe direkt an der Base.
- 15 Der Spaltungs- und Entfernungsschritt ist in jedem Zyklus vorhanden und muß unter milden Bedingungen (s.o.) verlaufen, so dass die Nukleinsäuren nicht beschädigt oder modifiziert werden.
- 20 Die Spaltung läuft bevorzugt chemisch (z.B. in milder saurer oder basischer Umgebung für eine Ester-Verbindung oder durch Zugabe eines Reduktionsmittel, z.B. Dithiothreitol oder Mercaptoethanol (Sigma) bei der Spaltung einer Disulfid-Verbindung), siehe Beispiel 1, oder physikalisch (z.B. durch  
25 Beleuchtung der Oberfläche mit Licht einer bestimmten Wellenlänge für die Spaltung einer photolabilen Gruppe, Dissertation „Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese" H. Giegrich, 1996, Konstanz) ab.
- 30 Nach der Spaltung verbleibt an der Base ein Linkerrest (A) (Fig.7c). Falls die nach der Spaltung am Linkerrest freigeordnete Mercapto-Gruppe weitere Reaktionen stört, kann sie mit bekannten Mitteln chemisch modifiziert werden (wie z.B.  
35 durch Disulfid- oder Iodacetatverbindungen).

Die Synthese eines spaltbaren Linkers wird an Beispielen gezeigt (vgl. Beispiele 1 und 2).

#### 4.5.7 Kombination von Polymerase und NT'

Insgesamt spielen die Größe, die Ladung und die chemische Struktur des Markers, die Länge des spaltbaren Linkers und des 5 Linker-Rests sowie auch die Wahl der Polymerase eine wichtige Rolle. Sie bestimmen gemeinsam, ob das markierte NT' durch die Polymerase in die wachsende Nukleinsäurekette eingebaut wird, und ob dadurch der Einbau des nächsten markierten NT' verhindert wird. Zwei Bedingungen sind dabei besonders zu berücksichtigen:

10

Einerseits ist es wichtig, dass die Polymerase die Nukleinsäurekette mit dem eingebauten modifizierten NT' nach der Spaltung des Linkers weiter verlängern kann. Es ist also wichtig, dass der Linkerrest "A" (Fig. 7c) nach der Spaltung 15 keine wesentliche Störung für die weitere Synthese darstellt. Andererseits müssen eingebaute, nicht gespaltene NTs' ein Hindernis darstellen. Es können viele für die Reaktion geeignete NTs' synthetisiert werden. Im einzelnen muß für jede Kombination aus Polymerase und NTs' eine Vorversuchsreihe 20 durchgeführt werden, in der die Tauglichkeit eines bestimmten NT'-Typs für die Sequenzierung erprobt wird.

Die Pufferbedingungen werden nach Angaben des Polymeraseherstellers gewählt. Die Reaktionstemperatur wird für 25 nicht thermostabile Polymerasen nach Angaben des Herstellers gewählt (z.B. 37°C für Sequenase Version 2), für thermostabile Polymerasen (z.B. Taq-Polymerase) beträgt die Reaktionstemperatur maximal den Temperaturwert (x). Dieser Temperaturwert (x) hängt von der T<sub>m</sub> des NSKF-Primer-Komplexes 30 und wird z.B. als T<sub>m</sub> (NSKF-Primer-Komplex) minus 5°C errechnet (z.B. T<sub>m</sub> ist 47°C, dann liegt die maximale Reaktionstemperatur bei 42°C). Diese Pufferbedingungen und Reaktionstemperatur werden weiter als "optimale Puffer- und Temperaturbedingungen" bezeichnet.

35

Die Reaktionszeit (entspricht der Dauer des Einbau-Schrittes in einem Zyklus) beträgt vorzugsweise weniger als eine Stunde,



- 36 -

idealerweise liegt die Reaktionszeit zwischen 10 sec und 10 min.

Als Beispiele von geeigneten Kombinationen zwischen NT' und 5 Polymerase sind folgende Kombinationen zu nennen:

a) NT' mit einem kurzen Linkerrest (Synthese siehe Beispiel 2, Fig. 7e,h,i): dNTP-SS-TRITC (L7), dNTP-SS-Cy3 (L11) in Kombination mit Sequenase Version 2, Klenow-Fragment DNA-Polymerase I 10 E.coli, Taq-Polymerase (GibcoBRL).

b) NT' mit einem langen Linkerrest (Synthese siehe Beispiel 1, Fig. 7f,g,j): dNTP-SS-TRITC (L14) in Kombination mit Sequenase Version 2 oder, Klenow-Fragment DNA-Polymerase I E.coli oder ProHATM-Polymerase (Eurogentech).

15

Die Tauglichkeit eines Linkerrests an der Base (A) für die Reaktion wird in einem Testsystem geprüft. Dabei werden gespaltene NTs' in eine Nukleinsäurekette nacheinander einbaut. Man verwendet z.B. dUTP' mit dem gewünschten gespaltenen 20 Linkerrest, poly-dA als Matrize, Oligo-dT20-Primer, die gewünschte Polymerase und führt unter für die jeweilige Polymerase geeigneten optimalen Puffer- und Temperaturbedingungen eine Reaktion durch. Die NT'-Konzentration liegt vorzugsweise zwischen 5 µmol/l und 200 µmol/l. Nach der 25 Reaktion wird die Anzahl der in die Nukleinsäurekette eingebauten NTs' analysiert, z.B. durch die Auftrennung der Länge nach in einem Gel. Für die Rückschlüsse auf die Tauglichkeit des Linkerrests kann man folgende Angaben verwenden: Wenn die Polymerase mehr als 20 NTs' einbauen kann, 30 so ist dieser Linkerrest für eine Sequenzierungsreaktion geeignet. Beim Einbau von weniger als 20 gespaltenen NTs' ist diese Kombination aus NT' und Polymerase nicht optimal für die Sequenzierungsreaktion.

35 Wenn ein passender Linkerrest feststeht, wird in einem weiteren Testsystem geprüft, ob die markierten, nicht gespaltenen NTs' als Semiterminatoren funktionieren. Das wird geprüft, indem die markierten NTs' unter für die Reaktion geeigneten optimalen



- 37 -

Puffer- und Temperaturbedingungen mit der Polymerase und einer Matrize inkubiert werden. Die NT'-Konzentration liegt vorzugsweise zwischen 5  $\mu\text{mol/l}$  und 200  $\mu\text{mol/l}$ . Die Matrize ist so zu wählen, dass der Einbau mehrerer NTs' nacheinander zu erwarten wäre, z.B. für dUTP' kann man polydA, wie im oben dargestellten Beispiel verwenden. Idealerweise baut die Polymerase nur ein einziges NT' ein.

Falls bei gegebenen optimalen Puffer- und Temperaturbedingungen durch eine Polymerase mehrere NTs' nacheinander eingebaut werden, kann man die Reaktionsparameter (z.B. NT'-Konzentration, Reaktionstemperatur) verändern und der jeweiligen Kombination aus Polymerase und NT' anpassen. Das wichtigste dabei ist, dass die Polymerase in der vorgegebenen Zeit (liegt vorzugsweise zwischen 10 sec und 10 min) ein zweites NT' nicht einbaut.

Erfindungsgemäß erfolgt diese Anpassung in einer Ausführungsform durch die Veränderung der Reaktionstemperatur. Die anderen Parameter der Reaktion werden dabei konstant gehalten.

Die NT'-Konzentration liegt bei diesen Experimenten üblicherweise zwischen 5  $\mu\text{mol/l}$  und 200  $\mu\text{mol/l}$ , vorzugsweise zwischen 10  $\mu\text{mol/l}$  und 100  $\mu\text{mol/l}$ . Die Konzentration der Polymerase und die Pufferbedingungen werden nach Angaben vom Hersteller gewählt. Die Dauer der Reaktion kann variieren und liegt vorzugsweise zwischen 10 sec und 10 min, was der Dauer des Einbau-Schrittes (a) in einem Zyklus entsprechen würde. Bei nicht thermostabilen Polymerasen wie z.B. Sequenase Version 2 (Amersham Pharmacia Biotech), exonuclease free Klenow-Fragment der DNA Polymerase I (Amersham Pharmacia Biotech) wird die Reaktionstemperatur von konventionellen 37°C vorzugsweise auf 20°C bis 30°C reduziert. Bei thermostabilen Polymerasen wie z.B. Taq-Polymerase (GibcoBRL), ProHATM-Polymerase (Eurogentech) wird die Reaktionstemperatur von konventionellen 70-75°C vorzugsweise auf Werte reduziert, die zwischen 30°C und dem Temperaturwert (x) liegen. Dieser Temperaturwert (x) hängt von der Tm des NSKF-Primer-

- 38 -

Komplexes und wird als  $T_m$  (NSKF-Primer-Komplex) minus  $5^{\circ}\text{C}$  errechnet (z.B.  $T_m$  ist  $47^{\circ}\text{C}$ , dann liegt der Temperaturwert (x) bei  $42^{\circ}\text{C}$ ).

5 In einer anderen bevorzugten Ausführungsform der Erfindung erfolgt die Anpassung der Reaktionsbedingungen durch die Verminderung der  $\text{NT}'$ -Konzentration auf unter  $5\text{ }\mu\text{mol/l}$ , die anderen Parameter der Reaktion (Pufferbedingungen, Temperaturbedingungen) werden konstant gehalten. Die  
10 Konzentration der  $\text{NT}'$  liegt vorzugsweise bei dieser Anpassung zwischen  $0.5\text{ }\mu\text{mol/l}$  und  $5\text{ }\mu\text{mol/l}$ . Die Dauer der Reaktion liegt zwischen 10 sec und 10 min. Das wichtigste bei der Wahl der  $\text{NT}'$ -Konzentration ist, dass die Polymerase in der vorgegebenen Zeit (liegt vorzugsweise zwischen 10 sec und 10 min) ein zweites  $\text{NT}'$   
15 nicht einbaut.

Nach Optimierung der Reaktionsbedingungen für den Einbau eines einzelnen  $\text{NT}'$  muß man die Reaktion mit gespaltenen  $\text{NTs}'$  wiederholen. Unter entsprechend geänderten Reaktionsparameter muß  
20 Polymerase die gespaltenen  $\text{NTs}'$  nacheinander einbauen können.

Die Optimierungsreaktion korreliert mit dem Einbauschritt, Schritt (b), in einem Zyklus. Die für die Optimierungsreaktion ermittelten Bedingungen, die Temperatur, die Konzentration an  
25  $\text{NT}'$ , die Pufferbedingungen, die Dauer der Reaktion werden für die Reaktion auf der Oberfläche übernommen.

Unter diesen Reaktionsbedingungen erfolgt der Einbau von  $\text{NT}'$  in die NSKF-Primer-Komplexe vorzugsweise so, dass an mehr als 50%  
30 der an der Sequenzierungsreaktion beteiligten NSKF-Primer-Komplexen in einem Zyklus ein markiertes  $\text{NT}'$  eingebaut wird, vorzugsweise an mehr als 90%. Das hängt damit zusammen, dass an manchen Nukleinsäureketten die Reaktion sehr langsam abläuft. Ein Einbau der  $\text{NTs}'$  an jeder komplementären Position in jedem  
35 Zyklus wird angestrebt, ist aber nicht erforderlich, weil nur die erfolgreichen Einbaureaktionen detektiert und ausgewertet werden; eine verzögerte Reaktion im Nachfolgenden Zyklus führt

- 39 -

nicht zu einem Sequenzierungsfehler.

Vorzugsweise wird für alle NTs' dieselbe Polymerase verwendet. Es können aber auch verschiedene Polymerasen für verschiedene 5 NTs' eingesetzt werden.

#### 4.5.8 Farbiges Kodierungsschema, Anzahl der Farbstoffe

Einen Zyklus kann man durchführen mit:

10

- a) vier verschieden markierten NT\*s
- b) zwei verschieden markierten NT\*s
- c) einem markierten NT\*
- d) zwei verschieden markierten NT\*s und zwei unmarkierten

15

NTs,

d.h.

a) Man kann alle 4 NTs mit verschiedenen Farbstoffen  
20 markieren und alle 4 gleichzeitig in die Reaktion einsetzen.  
Dabei erreicht man die Sequenzierung einer Nukleinsäurekette  
mit einer minimalen Anzahl von Zyklen. Diese Variante der  
Erfindung stellt allerdings hohe Anforderungen an das  
Detektionssystem: 4 verschiedene Farbstoffe müssen in jedem  
25 Zyklus identifiziert werden.

b) Zur Vereinfachung der Detektion kann eine Markierung mit  
zwei Farbstoffen gewählt werden. Dabei werden 2 Paare von NTs'  
gebildet, die jeweils verschieden markiert sind, z.B. A und G  
30 tragen die Markierung "X", C und U tragen die Markierung "Y".  
In die Reaktion in einem Zyklus (n) werden 2 unterschiedlich  
markierte NTs' gleichzeitig eingesetzt, z.B. C' in Kombination  
mit A', und im darauffolgenden Zyklus (n+1) werden dann U' und G'  
zugegeben.

35

c) Man kann auch nur einen einzigen Farbstoff zur  
Markierung aller 4 NTs' verwenden und pro Zyklus nur ein NT'  
einsetzen.

d) In einer technisch vereinfachten Ausführungsform werden pro Zyklus zwei unterschiedlich markierte NT's eingesetzt und zwei unmarkierte NTs (sogen. 2NT\*s / 2NTs-Methode). Diese Ausführungsform kann verwendet werden, um Varianten (z.B. Mutationen, oder alternativ gespleißte Gene) einer bereits bekannten Sequenz zu ermitteln.

#### 4.6 Detektionsapparatur

10

Einzelne Moleküle auf einer Oberfläche kann man mit verschiedenen Methoden untersuchen. Es sind mehrere Verfahren bekannt: z.B. AtomForce-Mikroskopie, Elektronen-Mikroskopie, Nahfeld-Fluoreszenz-Mikroskopie, Weitfeld-Fluoreszenz-Mikroskopie, TIR-Mikroskopie usw. (Science 1999 v.283 1667, Unger et al. BioTechniques 1999 v.27 S.1008, Ishijaima et al. Cell 1998 v.92 S.161, Dickson et al. Science 1996 v.274 S.966, Xie et al. Science 1994 v.265 S.361, Nie et al. Science 1994 v.266 S.1018, Betzig et al. Science 1993 v.262 S.1422).

20

Erfindungsgemäß werden Fluoreszenz-Signale einzelner in die Nukleinsäurekette eingebauter NTs' vorzugsweise mit einem Weitfeld-Fluoreszenz-Mikroskop (Epifluoreszenz) oder einem Laser-Scanning-Mikroskop (Epifluoreszenz) oder einem TIRF-Mikroskop (Total Internal Reflection Fluorescence Microscope).

Es sind verschiedene Varianten der Konstruktion einer solchen Apparatur möglich (Weston et al. J.Chem.Phys. 1998 v.109 S.7474, Trabesinger et al. Anal. Chem. 1999 v.71 S.279, Adachi et al. Journal of Microscopy 1999 v.195 S.125, Unger et al. BioTechniques 1999 v.27 S.1008, Ishijaima et al. Cell 1998 v.92 S.161, Dickson et al. Science 1996 v.274 S.966, Tokunaga et al. Bichem.Biophys.Res.Com. 1997 v.235 S.47, "Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995

- 41 -

J.Pawley Plenum Press ). Unterschiede in ihrem konkreten Aufbau ergeben sich aus der Variation ihrer Einzelteile. Die Vorrichtung für das Anregungslicht kann z.B. auf der Basis eines Lasers, einer Lampe oder von Dioden funktionieren. Für die Detektionsvorrichtung können sowohl CCD-Kameras als auch PMT dienen. Andere Beispiele für technische Details siehe ("Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press). Es ist nicht die Aufgabe dieser Erfindung, alle möglichen technischen Varianten einer Detektionsvorrichtung aufzuzählen. Der prinzipielle Aufbau einer geeigneten Apparatur wird in einem Schema Fig. 8 erläutert. Sie besteht aus folgenden Elementen:

- Lichtquelle zur Anregung der Fluoreszenz (1)
  - Lichtleitender Teil (2)
  - 20 Scantisch (3)
  - Vorrichtung zur Selektion von Spektren (4)
  - Detektionsvorrichtung (5)
  - Computer mit Steuerungs- und Analysefunktionen (6)
- 25 Diese Elemente der Apparatur können kommerziell erworben werden (Mikroskop-Firmen: Zeiss, Leica, Nikon. Olympus).

Im folgenden soll beispielsweise eine für die Detektion einzelner Moleküle geeignete Kombination aus diesen Elementen  
30 vorgestellt werden:

- Weitfeld-Fluoreszenz-Mikroskop Axioplan 2 (Zeiss) mit Quecksilberdampflampe
- Objektiv Planneofluar 100x, NA 1.4 (Zeiss)
- 35 Kamera Photometrix oder AxioCam (Zeiss)
- Computer mit Software zur Steuerung und Analyse

Nachfolgend soll die Vorgehensweise bei der Detektion erläutert werden. Man beachte dabei die allgemeinen Regeln der Fluoreszenzmikroskopie ("Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press).

10 Die Detektion umfaßt folgende Phasen:

- 1) Vorbereitung zur Detektion
- 2) Durchführung eines Detektionsschrittes in jedem Zyklus, wobei jeder Detektionsschritt als Scanvorgang abläuft und  
15 folgende Operationen umfaßt:
  - a) Einstellung der Position des Objektivs (X,Y-Achse),
  - b) Einstellung der Fokusebene (Z-Achse),
  - c) Detektion der Signale einzelner Moleküle, Zuordnung des Signals zu NT' und Zuordnung des Signals zum  
20 jeweiligen NSKF,
  - d) Verschiebung zur nächsten Position auf der Oberfläche.

Die Signale von in die NSKFs eingebauten NTs' werden durch das Abscannen der Oberfläche registriert. Der Scanvorgang kann in  
25 verschiedener Weise ausgeführt werden ("Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of  
30 biological confocal microscopy" 1995 J.Pawley Plenum Press). Beispielsweise wird ein diskontinuierlicher Scanvorgang gewählt. Dabei wird das Objektiv schrittweise über die Oberfläche bewegt (Fig. 8a), so dass von jeder Oberflächenposition ein zweidimensionales Bild (2D-Bild)  
35 entsteht (Fig 8b,c), für Versuchsanordnung s. Beispiel 6.

Dieses 2D-Bild kann mit verschiedenen Methoden erstellt werden: z.B. durch den Laser-Scan einer Position des Mikroskopfeldes



(Laser-Scanning-Mikroskopie) oder durch eine Kameraaufnahme an einer Position (vgl. Handbücher der Mikroskopie). Als Beispiel wird die Detektion einzelner Moleküle mit einer CCD-Kamera beschrieben.

5

Die Detektion wird schematisch am Beispiel der Sequenzierung eines 1Mb langen DNA-Stücks erläutert:

10 1) Vorbereitung zur Detektion:

Am Anfang wird festgelegt, wie viele NSKF-Sequenzen zur Rekonstruktion der ursprünglichen Sequenz analysiert werden müssen. Im Fall einer Rekonstruktion nach dem Schrotschuß-Verfahren ("Automated DNA sequencing and analysis" S. 231 ff. 15 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257) spielen folgende Faktoren eine Rolle: 1) Von jedem NSKF wird bei der Sequenzierung eine Sequenz von ca. 300-500 NTs 20 bestimmt. 2) Die Gesamtlänge der zu analysierenden Sequenz ist wichtig. 3) Bei der Sequenzierung muß ein bestimmtes Maß an Redundanz erreicht werden, um die Genauigkeit zu steigern und eventuelle Fehler zu korrigieren. Insgesamt ist zur Rekonstruktion des größten Teils der ursprünglichen Sequenz die 25 etwa 10- bis 100-fache Menge an Rohsequenzen erforderlich, d.h. bei diesem Beispiel mit einer Mb, werden 10 bis 100 Mb Rohsequenzdaten gebraucht. Bei einer durchschnittlichen Sequenzlänge von 400 bp pro NSKF benötigt man entsprechend 25.000 bis 250.000 DNA-Fragmente.

30

2) Durchführung eines Detektionsschrittes in jedem Zyklus

Zur Sequenzierung müssen die Positionen der NSKFs bestimmt werden, damit man eine Grundlage für die Zuordnung der Signale hat. Die Kenntnis dieser Positionen erlaubt eine Aussage 35 darüber, ob die Signale einzelner Moleküle von eingebauten NTs' stammen oder von zufällig an die Oberfläche gebundenen NTs'. Diese Positionen können mit verschiedenen Methoden identifiziert werden.

In einer bevorzugten Ausführungsform werden die Positionen gebundener NSKF-Primer-Komplexe während der Sequenzierung identifiziert. Dabei wird die Tatsache genutzt, dass die  
5 Signale von den in die Nukleinsäurekette eingebauten NTs' immer dieselben Koordinaten haben. Das ist durch die Fixierung der Nukleinsäureketten gewährleistet. Die unspezifisch gebundenen NTs' binden zufällig an verschiedenen Stellen der Oberfläche.

10 Zur Identifizierung der Positionen von fixierten NSKFs werden die Signale auf Übereinstimmung ihrer Koordinaten aus mehreren aufeinander folgenden Zyklen überprüft. Das kann z.B. am Anfang der Sequenzierung erfolgen. Die übereinstimmende Koordinaten werden als Koordinaten der DNA-Fragmente bewertet und gespei-  
15 chert.

Das Scan-System muß reproduzierbar über mehrere Zyklen die Oberfläche absannen können. X,Y und Z-Achsen-Einstellungen an jeder Oberflächenposition können von einem Computer  
20 kontrolliert werden. Stabilität und Reproduzierbarkeit der Einstellung von Objektivpositionen in jedem Scanvorgang entscheiden über die Qualität der Detektion und somit über die Identifizierung der Signale einzelner Moleküle.

25 a) Einstellung der Position des Objektivs (X,Y-Achse)

Die mechanische Instabilität der kommerziell erhältlichen Scantische und die geringe Reproduzierbarkeit der wiederholten Einstellung derselben X,Y-Positionen machen eine präzise Analysen  
30 der Signale einzelner Moleküle über mehrere Zyklen schwierig. Es existieren viele Möglichkeiten, eine Übereinstimmung der Koordinaten bei wiederholten Einstellungen zu verbessern bzw. mögliche Abweichungen zu kontrollieren. Als Beispiel wird eine Kontrollmöglichkeit angeführt. Nach einer groben mechanischen  
35 Einstellung der Objektivposition wird ein Kontrollbild von einem mit der Oberfläche fest verbundenen Muster gemacht. Auch wenn die mechanische Einstellung nicht exakt dieselben Koordinaten aufweist (Abweichungen bis zu 10 µm sind durchaus

- 45 -

möglich), kann man mittels optischer Kontrolle eine Korrektur vornehmen. Das Kontrollbild vom Muster dient als Koordinatensystem für das Bild mit Signalen von eingebauten NTs'. Eine Voraussetzung für eine solche Korrektur ist, dass  
5 keine weiteren Bewegungen der Oberfläche zwischen diesen beiden Aufnahmen gemacht werden. Signale von einzelnen Molekülen werden in Relation zum Muster gesetzt, so dass eine X,Y-Abweichung in der Musterposition gleiche X,Y-Abweichung in der Position der Signale einzelner Moleküle bedeutet. Das  
10 Kontrollbild vom Muster kann vor, während oder nach der Detektion einzelner Moleküle gemacht werden. Ein solches Kontrollbild muß entsprechend bei jeder Einstellung auf einer neuen Oberflächenposition gemacht werden.

15 b) Einstellung der Fokusebene (Z-Achse)

Die Oberfläche ist nicht absolut plan und weist verschiedene Unebenheiten auf. Dadurch verändert sich der Oberfläche-Objektiv-Abstand beim abscannen benachbarter Stellen. Diese Unter-  
20 schiede im Abstand können dazu führen, dass einzelne Moleküle die Fokusebene verlassen und so der Detektion entgehen.

Aus diesem Grund ist es wichtig, dass beim Abscannen der Oberfläche eine reproduzierbare Einstellung der Fokusebene an jeder  
25 Objektivposition erreicht wird.

Es gibt verschiedene Möglichkeiten, die Fokusebene reproduzierbar einzustellen. Beispielsweise kann folgende Methode angewendet werden: Da die Anregung einzelner Moleküle zum  
30 Auslösen ihrer Fluoreszenz führen kann, wird auf die Oberfläche ein Marker aufgebracht, der zur Einstellung der Fokusebene dient. Danach erfolgt die Detektion der Signale einzelner Moleküle. Der Marker kann beliebiger Natur sein (z.B. Farbstoff oder Muster), darf aber die Detektion und die  
35 Reaktion nicht beeinträchtigen.

c) Detektion der Signale einzelner Moleküle, Zuordnung des Signals zu NT' und Zuordnung des Signals zum jeweiligen NSKF.

Das mit Hilfe des Detektionssystems erzeugte zweidimensionale Bild der Reaktionsoberfläche enthält die Signalinformationen von in die NSKFs eingebauten NT's. Diese müssen vor der weiteren  
 5 Verarbeitung aus der Gesamtdatenmenge der Bildinformationen mit geeigneten Methoden extrahiert werden. Die dazu notwendigen Algorithmen zur Skalierung, Transformation und Filterung der Bildinformationen zählen zum Standardrepertoire der digitalen Bildverarbeitung und Mustererkennung (Haberäcker P. "Praxis der  
 10 Digitalen Bildverarbeitung und Mustererkennung". Hanser-Verlag, München, Wien, 1995; Galbiati L.J. "Machine vision and digital image processing fundamentals". Prentice Hall, Englewood Cliffs, New Jersey, 1990). Die Signalextraktion erfolgt vorzugsweise über ein Grauwertbild, das die  
 15 Helligkeitsverteilung der Reaktionsoberfläche für den jeweiligen Fluoreszenzkanal abbildet. Wenn bei der Sequenzierungsreaktion mehrere Nukleotide mit unterschiedlichen Fluoreszenz-Farbstoffen verwendet werden, kann zunächst für jedes verwendete fluoreszenzmarkierte Nukleotid (A,T,C,G oder  
 20 U) ein separates Grauwert-Bild erzeugt werden. Dafür können prinzipiell 2 Verfahren angewendet werden:

1. Durch Verwendung von geeigneten Filtern (Zeiss-Filtersätze) wird für jeden Fluoreszenzkanal ein Grauwertbild erzeugt.

25

2. Aus einem aufgenommenen Mehrkanal-Farb-Bild werden mit Hilfe eines geeigneten Algorithmus durch ein Bildverarbeitungsprogramm die relevanten Farbkanäle extrahiert und jeweils als Grauwertbild einzeln weiterverarbeitet. Zur  
 30 Kanalextraktion wird dabei ein für den jeweiligen Kanal spezifischer Farb-Schwellwertalgorithmus eingesetzt. So entstehen zunächst aus einem Mehrkanal-Farbbild einzelne Grauwertbilder 1 bis N. Diese Bilder definieren sich wie folgt:

35  $GB_N = (s(x,y))$  einkanaliges Grauwertbild  
 $N = \{1, \dots, \text{Anzahl der Fluoreszenzkanäle}\}$ .  
 $M = \{0, 1, \dots, 255\}$  Grauwertmenge  
 $S = (s(x,y))$  Bildmatrix des Grauwertbildes

- 47 -

$x=0,1,\dots,L-1$       Bildzeilen  
 $y=0,1,\dots,R-1$       Bildspalten  
 $(x,y)$               Ortskoordinaten eines Bildpunktes  
 $s(x,y) \in M$           Grauwert des Bildpunktes.

5

Aus dieser Datenmenge wird nun durch ein geeignetes Programm die relevante Bildinformation extrahiert. Ein solches Programm sollte folgende Arbeitsschritte realisieren:

10 Für  $GB_1$  bis  $GB_N$  durchführen:

I. Vorverarbeitung des Bildes, so zum Beispiel gegebenenfalls Reduktion des durch die Digitalisierung der Bildinformation entstandenen Bildrauschens, etwa durch Grauwertglättung.

15

II. Prüfung jedes Bildpunkt  $(x,y)$  des Grauwertbildes, ob dieser Punkt im Zusammenhang mit den ihn umgebenden unmittelbaren und weiter entfernten Nachbarbildpunkten die Eigenschaften eines Fluoreszenzpunktes erfüllt. Diese Eigenschaften hängen unter  
20 anderem von der verwendeten Detektionsapparatur und der Auflösung des Grauwertbildes ab. Sie können beispielsweise ein typisches Verteilungsmuster von Helligkeits-Intensitätswerten über einer den Bildpunkt umgebenden Matrix darstellen. Die dazu verwendbaren Methoden der Bildsegmentierung reichen von  
25 einfachen Schwellwertverfahren bis hin zur Verwendung neuronaler Netze.

Erfüllt ein Bildpunkt  $(x,y)$  diese Anforderungen, dann folgt ein Vergleich mit den Koordinaten von in bisher durchgeführten  
30 Sequenzierungszyklen identifizierten NSKFs. Bei einer Übereinstimmung erfolgt die Zuordnung des Signals mit dem aus dem jeweiligen Fluoreszenzkanal hervorgehenden Nukleotid zu diesem NSKF. Signale mit nicht übereinstimmenden Koordinaten werden als Hintergrundsignale bewertet und verworfen. Die Analyse der  
35 Signale kann parallel zum Scanvorgang erfolgen.

In einer beispielhaften Ausführung wurde ein 8-Bit-Grauwertbild mit einer Auflösung von 1317 x 1035 Pixel verwendet. Um die



- 48 -

durch die Digitalisierung entstandenen Veränderungen am Bild zu reduzieren, erfolgte zunächst eine Vorverarbeitung des Gesamtbildes: Jedem Bildpunkt wurde der Mittelwert der Helligkeiten seiner 8-Nachbarn zugewiesen. Bei der gewählten Auflösung 5 entsteht dadurch ein für einen Fluoreszenzpunkt typisches Muster eines zentralen Bildpunkt mit dem größten Helligkeitswert und Nachbarbildpunkten mit nach allen Seiten hin abfallenden Helligkeiten. Erfüllte ein Bildpunkt diese Kriterien und überschritt der zentrifugale Helligkeitsabfall 10 einen bestimmten Schwellenwert (zur Exklusion zu schwacher Fluoreszenzpunkte), dann wurde dieser zentrale Bildpunkt als Koordinate eines Fluoreszenzpunktes gewertet.

d) Verschiebung des Objektivs zur nächsten Position auf der 15 Oberfläche. Nach der Detektion der Signale einzelner Moleküle wird das Objektiv über einer anderen Position der Oberfläche positioniert.

Insgesamt kann beispielsweise eine Folge von Aufnahmen mit der 20 Kontrolle der X,Y-Position, der Einstellung der Fokusebene und mit der Detektion einzelner Moleküle bei jeder neuen Objektivposition gemacht werden. Diese Schritte können durch einen Computer gesteuert werden.

#### 25 4.7 Zeitlicher Ablauf der Verfahrensschritte

Der Scanvorgang sowie die biochemische Reaktion nehmen eine gewisse Zeit in Anspruch. Wenn man diese Vorgänge nacheinander schaltet, kann man eine optimale Leistung der Apparatur errei- 30 chen. In einer bevorzugten Ausführung wird die Reaktion auf zwei getrennten Oberflächen durchgeführt (Fig. 9).

Als Beispiel kann eine Oberfläche mit gebundenen NSKF-Primer-Komplexen in 2 räumlich isolierte Teile getrennt werden, so 35 dass Reaktionen auf diesen beiden Teilen unabhängig voneinander ablaufen können. In einem anderen Beispiel können NSKFs auch von vornherein auf 2 getrennten Oberflächen immobilisiert werden.



Danach wird die Reaktion gestartet. Das Prinzip dabei ist, dass während auf einem Teil der Oberfläche die Reaktions- und Waschschr

5 kritte ablaufen, der zweite Teil abgescannt wird. Dadurch kann man einen kontinuierlichen Ablauf der Analyse erreichen und die Geschwindigkeit der Sequenzierung steigern.

Die Anzahl der Oberflächen, auf denen die Reaktion abläuft, kann auch größer als 2 sein. Das erscheint dann sinnvoll, wenn

10 die Reaktion als zeitlich limitierender Schritt auftritt, d.h. die Detektion der Signale auf der Oberfläche schneller als die Reaktions- und Waschschr

15 itte abläuft. Um die Gesamtdauer der Reaktion an die Detektionsdauer anzupassen, kann jeder einzelne Schritt der Reaktion auf einer einzelnen Oberfläche mit einer zeitlichen Verzögerung im Vergleich zur nächsten Oberfläche ablaufen.

Die Erfindung wird nachfolgend anhand von Beispielen verdeutlicht.

20

### Beispiele

#### Beispiel 1:

25

Modifiziertes dUTP mit einem langen spaltbaren Linker (Fig. 7f-1) Als Ausgangssubstanzen dienen 5-(3-Aminoallyl)-2'-deoxyuridin- 5'-triphosphat, AA-dUTP, (Sigma) , 3,3'-Dithio-

30 bis(propionsäure- N-Hydroxysuccinimidester), DTBP-NHS, (Sigma), 2-Mercaptoethylamin, MEA, (Sigma). Zu 100 µl 50mmol/l Lösung von AA-dUTP in 100mmol/l Borat-Puffer pH 8.5 werden 3 Äquivalente an DTBP-NHS in DMF (25 µl 0.4mol/l Lösung) zugegeben. Das Reaktionsgemisch wird 4 Stunden bei RT. inkubiert. Anschließend wird konz. Ammoniumacetat-Lösung (pH 9) zugegeben bis die

35 Gesamtkonzentration an  $\text{CH}_3\text{COONH}_4$  in der Reaktionslösung 100mmol/l ist, und die Reaktion wird eine weitere Stunde inkubiert. Danach werden zu diesem Gemisch 200 µl 1mol/l MEA-Lösung, pH 9, zugegeben und eine Stunde bei RT inkubiert.

- 50 -

Anschließend wird zu diesem Gemisch solange eine gesättigte Lösung an I<sub>2</sub> in 0.3mol/l KI-Lösung zugetropft, bis die Iodfarbe bestehen bleibt. Die modifizierten Nukleotide werden auf einer DEAE-Cellulose-Säule in Ammoniumcarbonat-Gradient (pH 8.5) von 5 anderen Reaktionsprodukten abgetrennt. Isolierung des Nukleotids mit dem spaltbaren Linker erfolgt auf RP-HPLC. An diesen Linker können nun Farbstoffe mit verschiedenen Methoden gekoppelt werden ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes, Waggoner 10 Method in Enzymology 1995 v.246, S.362, Jameson et al. Method in Enzymology 1997, v.278, S.363).

Auch andere Nukleotidanaloga (z.B. nach Hobbs et al. US Patent 5,047,519, Khan et al. US Patent 5,821,356) können in die 15 Reaktion eingesetzt werden, so dass Nukleotidanaloga mit Strukturen in Fig. 7f-2,3,4 und 7 g-1,2 erzeugt werden können.

Als Beispiel der Ankopplung eines Farbstoffs an den Linker wird die Ankopplung von TRITC (Tetramethylrhodamin-isothiocyanat) 20 angegeben (NT'-Struktur Fig. 7j)

Das mit dem spaltbaren Linker modifizierte dNTP (300 nmol) wird in 30 µl 100mmol/l Natrium-Borat-Puffer pH 9 aufgelöst (10mmol/l NT'). Dazu werden 10 µl 10mmol/l TRITC in 25 Dimethylformamid (DMF) gegeben und 4h bei RT inkubiert. Die Reinigung des mit dem Farbstoff modifizierten NT' erfolgt über RP-HPLC in einem Methanol-Wasser Gradienten.

Das so hergestellte NT\* erfüllt die Anforderungen des Einbaus 30 in den DNA-Strang, des Fluoreszenznachweises und Kettenabbruchs nach dem Einbau und der Aufhebung der Hemmung, die für das Gelingen des erfindungsgemäßen Verfahrens notwendig sind.

35 Beispiel der Spaltung der Disulfidverbindung im modifizierten NT'. Die Spaltung erfolgt durch Zugabe von 20 bis 50mmol/l Dithiothreitol-Lösung (DTT) oder Mercaptoethanol-Lösung (Sigma), pH 8, auf die Reaktionsoberfläche. Die Oberfläche wird

10 min. mit dieser Lösung inkubiert, danach wird die Lösung entfernt und die Oberfläche mit einer Pufferlösung zur Entfernung von DTT- bzw. Mercaptoethanol-Resten gewaschen.

5

Beispiel 2:

Modifiziertes dUTP ( $\text{dUTP-SS-CH}_2\text{CH}_2\text{NH}_2$ ) mit einem kurzen spaltbaren Linker (Fig. 7e-1). Als Ausgangssubstanzen dienen:  
10 Bis-dUTP, synthetisiert nach Hanna (Method in Enzymology 1989, v.180, S.383), 2-Mercaptoethylamin (MEA) (Sigma).

Zu 400  $\mu\text{l}$  100mmol/l Bis-dUTP in 40mmol/l Boratpuffer pH 8.5 werden 100 $\mu\text{l}$  100mmol/l MEA-Lösung pH 8.5 in  $\text{H}_2\text{O}$  zugegeben und 1  
15 Stunde bei RT inkubiert. Anschließend wird zu diesem Gemisch solange eine gesättigte Lösung an  $\text{I}_2$  in 0.3mol/l KI-Lösung zuge-  
tropft, bis die Iodfarbe bestehen bleibt. Die Nukleotide (Bis-  
dUTP und  $\text{dUTP-SS-CH}_2\text{CH}_2\text{NH}_2$ ) können z.B. durch eine Ethanol-  
Präzipitation oder auf einer DEAE-Cellulose-Säule im  
20 Ammoniumcarbonat-Gradienten (pH 8.5) von anderen  
Reaktionsprodukten abgetrennt werden. Bis-dUTP stört bei der  
anschließenden Ankopplung eines Farbstoffs an die Aminogruppe  
des Linkers nicht, so dass die Abtrennung des  $\text{dUTP-SS-CH}_2\text{CH}_2\text{NH}_2$   
von bis-dUTP im Endreinigungsschritt erfolgen kann.

25

In einer ähnlichen Weise kann auch dCTP (Fig.7-e2) modifiziert werden, dabei dient Bis-dCTP als Ausgangssubstanz (synthetisiert nach Hanna et al. Nucleic Acid Research 1993, v.21, S.2073).

30

An den Linker können nun Farbstoffe mit verschiedenen Methoden gekoppelt werden ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes, Waggoner Method in Enzymology 1995 v.246, S.362, Jameson et al. Method  
35 in Enzymology 1997, v.278, S.363).

Als Beispiel der Ankopplung eines Farbstoffs an den Linker wird die Ankopplung des FluoroLink™ Cy3 monofunktional dye

(Amersham Pharmacia biotech) (NT'-Struktur Fig. 7i) angegeben. Das ist ein monofunktionaler NHS-Ester-Fluoreszenzfarbstoff. Die Reaktion wird nach Angaben des Herstellers durchgeführt:

Das mit dem spaltbaren Linker modifizierte dNTP (300 nmol) wird 5 in 300 µl 100mmol/l Natrium-Borat-Puffer pH 8.5 aufgelöst. Dazu wird Farbstoff (300nmol) gegeben und 1h bei RT inkubiert. Die Reinigung des mit dem Farbstoff modifizierten NT' erfolgt über RP-HPLC in einem Methanol-Wasser Gradienten.

10 Ein weiteres Beispiel der Ankopplung eines Farbstoffs an den Linker wird die Ankopplung von TRITC (Tetramethylrhodamin-5-isothiocyanat, Molecular Probes) angegeben (dUTP-SS-TRITC Fig.7h).

Das mit dem spaltbaren Linker modifizierte dNTP (300 nmol) wird 15 in 30 µl 100mmol/l Natrium-Borat-Puffer pH 9 aufgelöst (10mmol/l NT'). Dazu werden 10 µl 10mmol/l TRITC in DMF gegeben und 4h bei RT inkubiert. Die Reinigung des mit dem Farbstoff modifizierten NT' erfolgt über RP-HPLC in einem Methanol-Wasser Gradienten.

20 Das so hergestellte NT\* erfüllt die Anforderungen des Einbaus in den DNA-Strang, des Fluoreszenznachweises und Kettenabbruchs nach dem Einbau und der Aufhebung der Hemmung, die für das Gelingen des erfindungsgemäßen Verfahrens notwendig sind.

25

Beispiel der Spaltung der Disulfidverbindung im modifizierten NT'. Die Spaltung erfolgt durch Zugabe von 20 bis 50mmol/l Dithiothreitol-Lösung (DTT) oder Mercaptoethanol-Lösung (Sigma), pH 8, auf die Reaktionsoberfläche. Die Oberfläche 30 wird 10 min. mit dieser Lösung inkubiert, danach wird die Lösung entfernt und die Oberfläche mit einer Pufferlösung zur Entfernung von DTT- bzw. Mercaptoethanol-Resten gewaschen.

Weitere NT\*s, wie z.B. in Fig. 7k, 7l, 7m dargestellt, können 35 ähnlich zu den in Beispielen 1 und 2 beschriebenen NT-Analoga synthetisiert und im Verfahren eingesetzt werden. Für die einzelnen Syntheseschritte siehe beispielsweise J.L.Ruth et al. Molecular Pharmacology 1981 v.20 S.415, L. Ötvös et al. NAR

- 53 -

1987 v.15 S.1763, G.E.Wright et al. Pharmac Ther. 1990 v.47, S.447, „Nucleotide Analogs; Synthesis and Biological Function“ K.H. Scheit 1980, Wiley-Interscience Publication, "Nucleic acid chemistry" Ed. L.B.Townsend, v.1-4, Wiley-Interscience Publication, "Chemistry of Nucleosides and Nucleotides" Ed. L.B.Townsend, v.1-3, Plenum Press.

Beispiel 3:

10

Sequenzanalyse mit 4 markierten NTs'

Bei einer bevorzugten Ausführungsform der Erfindung werden alle vier in die Reaktion eingesetzten NTs' mit Fluoreszenz-15 farbstoffen markiert.

3A. Rekonstruktion der ursprünglichen Sequenzen nach dem Schrotschuß-Prinzip ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 20 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). (Dieses Prinzip ist insbesondere bei der Analyse neuer, unbekannter Sequenzen geeignet.)

25

3A-1Sequenzierung eines langen DNA-Stücks (Fig.1)

Im folgenden soll anhand der Sequenzierung eines 1Mb langen DNA-Stückes schematisch die Sequenzierung langer 30 Nukleinsäureketten dargestellt werden (Fig. 1). Der Sequenzierung liegt das Shotgun-Prinzip zugrunde ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. 35 NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Das zu analysierende Material wird für die Sequenzierungsreaktion vorbereitet, indem es in Fragmente von vorzugsweise 50 bis 1000 bp Länge zerlegt wird. Jedes

- 54 -

Fragment wird anschließend mit einer Primerbindungsstelle und einem Primer versehen. Dieses Gemisch aus verschiedenen DNA-Fragmenten wird nun auf einer planen Oberfläche fixiert. Die nicht gebundenen DNA-Fragmente werden durch  
5 einen Waschschrift entfernt. Danach wird die Sequenzierungsreaktion an der gesamten Reaktionsoberfläche durchgeführt. Zur Rekonstruktion einer 1 Mb langen DNA-Sequenz sollten die Sequenzen von NSKFs vorzugsweise länger als 300 NTs sein, durchschnittlich ca. 400 bp. Da pro  
10 Zyklus nur jeweils ein markiertes NT\* eingebaut wird, sind mindestens 400 Zyklen zur Sequenzierung notwendig.

Insgesamt ist zur Rekonstruktion der ursprünglichen Sequenz die etwa 10- bis 100-fache Menge an Rohsequenzen erforderlich,  
15 d.h. 10 bis 100 Mb. Bei einer durchschnittlichen Sequenzlänge von ca. 400 bp pro NSKF benötigt man entsprechend 25.000 bis 250.000 DNA-Fragmente, um mehr als 99,995% der Gesamtsequenz abzudecken.

20 Die ermittelten NSKF-Sequenzen stellen eine Population von überlappenden Teilsequenzen dar, die sich mit kommerziell erhältlichen Programmen zur Gesamtsequenz der NSK zusammenfügen lassen ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press , Huang et al.  
25 Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257).

### 3A-2Sequenzierung der Genprodukte am Beispiel der cDNA-Sequenzierung 30 zierung

In einer bevorzugten Ausführungsform können statt einer Sequenz mehrere Sequenzen in einem Ansatz analysiert werden. Die ursprünglichen Sequenzen können aus den  
35 gewonnen Rohdaten z.B. nach dem Schrotschuß-Prinzip rekonstruiert werden.

Zunächst werden NSKFs erzeugt. Man kann z.B. mRNA in eine



doppelsträngige cDNA überführen und diese cDNA mit Ultraschall fragmentieren. Anschließend werden diese NSKFs mit einer Primerbindungsstelle versehen, denaturiert, immobilisiert und mit einem Primer hybridisiert. Zu beachten ist bei dieser Variante der Probenvorbereitung, dass die cDNA-Moleküle unvollständige mRNA-Sequenzen darstellen können (Method in Enzymology 1999, v.303, S.19 und andere Artikel in diesem Band, "cDNA library protocols" 1997 Humana Press).

Eine andere Möglichkeit bei der Generierung einzelsträngiger NSKFs von mRNA besteht in der reversen Transkription der mRNA mit randomisierten Primern. Dabei werden viele relativ kurze antisense DNA-Fragmente gebildet (Zhang-J et al. Biochem.J. 1999 v.337 S.231, Ledbetter et al. J.Biol.Chem. 1994 v.269 S.31544, Kolls et al. Anal.Biochem. 1993 v.208 S.264, Decraene et al. Biotechniques 1999 v.27 S.962). Diese Fragmente können anschließend mit einer Primerbindungsstelle versehen werden (s.o). Weitere Schritte entsprechen oben beschriebenen Vorgängen. Mit dieser Methode können komplette mRNA-Sequenzen (vom 5'- bis zum 3'-Ende) analysiert werden, da die randomisierten Primer über die gesamte Länge der mRNA binden.

Immobilisierte NSKFs werden mit einer der oben angeführten Ausführungsformen der Sequenzierung analysiert. Da mRNA-Sequenzen wesentlich weniger repetitive Sequenzen aufweisen als z.B. genomische DNA, kann die Anzahl der detektierten Signale der eingebauten NTs' von einem NSKF geringer als 300 sein und liegt vorzugsweise zwischen 20 und 1000. Die Anzahl der NSKFs, die analysiert werden müssen, errechnet sich nach denselben Prinzipien wie bei einer Schrotschuß-Rekonstruktion einer langen Sequenz.

Aus NSKF-Sequenzen werden nach den Prinzipien des Schrotschuß-Verfahrens die ursprünglichen Gensequenzen rekonstruiert.

Diese Methode erlaubt die gleichzeitige Sequenzierung von vielen mRNAs ohne vorherige Klonierung.

### 5 3B. Analyse von Sequenzvarianten

Die Bestätigung einer bereits bekannten Sequenz oder der Nachweis von Varianten dieser Sequenz stellt sehr viel geringere Ansprüche an die Länge und Redundanz der  
10 ermittelten NSKF-Sequenzen. Auch die Sequenzbearbeitung ist in diesem Fall einfacher. Die Vollsequenz braucht nicht neu rekonstruiert zu werden. Die NSKF-Sequenzen werden vielmehr mit Hilfe eines kommerziell erhältlichen Programms der Vollsequenz zugeordnet und eventuelle Abweichungen  
15 detektiert. Einem solchen Programm kann z.B. BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to computational Biology" 1995 M.S. Waterman Chapman & Hall).

Die zu analysierende Sequenz wird mit einer der oben  
20 genannten Methoden in NSKFs überführt. Diese NSKFs werden mit dem erfindungsgemäßen Verfahren sequenziert, wobei man sowohl einen einheitlichen Primer und eine einheitliche Primerbindungsstelle als auch unterschiedliche, sequenzspezifische Primer und natürliche, in der zu  
25 untersuchenden Gesamtsequenz vorkommende Primerbindungsstellen, s. Beispiel 5, verwenden kann. Anschließend werden die ermittelten Sequenzen von NSKFs nicht nach dem Schrotschuß-Verfahren zusammengesetzt, sondern mit der Referenzsequenz verglichen und auf diese  
30 Weise ihren Positionen in der Vollsequenz zugeordnet. Dabei kann es sich um genomische oder cDNA-Sequenzen handeln.

Im Gegensatz zu einer Rekonstruktion nach dem Schrotschuß-Verfahren braucht man für die Analyse einer Sequenzvariante  
35 erheblich weniger Rohsequenzdaten. So kann die 5- bis 10-fache Rohsequenzmenge ausreichend für die Wiederherstellung einer neuen Variante einer Vollsequenz sein. Mit dem Schrotschuß-Verfahren wird für eine Wiederherstellung eine

10- bis 100-fache Menge an Rohsequenzen benötigt  
("Automated DNA sequencing and analysis" S. 231 ff. 1994 M.  
Adams et al. Academic Press, Huang et al. Genom Res. 1999  
v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al.  
5 NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1  
S.257).

Die Länge der ermittelten NSKF-Sequenzen soll für eine  
eindeutige Zuordnung zu einer bestimmten Position in der  
10 Referenzsequenz ausreichend sein, so können z.B bereits  
Sequenzen mit einer Länge von 20 NTs (z.B. aus nicht  
repetitiven Abschnitten im menschlichen Genom) eindeutig  
identifiziert werden. Für die Vergleichsanalyse der repet-  
itiven Abschnitte werden längere Sequenzen benötigt. Die  
15 genaue Länge der Sequenzen hängt dabei von der Aufgaben-  
stellung ab. Vorzugsweise beträgt die Länge der ermittelten  
NSKF-Sequenzen bei der Analyse von nicht repetitiven  
Abschnitten mehr als 20 NTs. Für die Analyse der  
repetitiven Abschnitte liegt sie vorzugsweise über 500 NTs.

20 Die Zielsetzungen bei der Sequenzierung neuer Varianten  
einer bereits bekannten Vollsequenz können sehr unter-  
schiedlich sein. Meist wird ein Vergleich der neu ermittel-  
ten Sequenz mit der bekannten Vollsequenz/Referenzsequenz  
25 angestrebt. Dabei können die beiden Sequenzen aus evolutio-  
när unterschiedlich weit auseinanderliegenden Spezies  
stammen. Verschiedene Parameter der Zusammensetzung dieser  
beiden Sequenzen können verglichen werden. Als Beispiele  
für eine solche Analyse dienen: Mutations- oder  
30 Polymorphismusanalysen und die Analyse von alternativ ge-  
speicherten Genprodukten.

Nachfolgend soll schematisch und beispielhaft ein Vergleich  
der zu untersuchenden Sequenz mit einer Referenzsequenz  
35 ohne vorherige Rekonstruktion der zu analysierenden Sequenz  
betrachtet werden. Ein solcher Vergleich kann z.B. zur  
Mutations- oder SNP-Analyse dienen.

3B-1

Eine lange, zu analysierende Sequenz, z.B. 1 Mb, wird in  
NSKFs mit einer der oben genannten Methode geteilt. Diese  
5 NSKFs werden unter Verwendung einheitlicher Primer mit dem  
erfindungsgemäßen Verfahren sequenziert. Die ermittelten  
Sequenzen von jedem einzelnen NSKF werden direkt mit der  
Referenzsequenz verglichen. Die Referenzsequenz dient dabei  
als Grundlage für die Zuordnung ermittelter NSKF-Sequenzen,  
10 so dass die aufwendige Rekonstruktion nach dem Schrotschuß-  
Verfahren entfällt. Vorzugsweise beträgt die Länge der  
ermittelten NSKF-Sequenzen bei der Analyse von nicht-  
repetitiven Abschnitten mehr als 20 NTs. Für die Analyse  
der repetitiven Abschnitte liegt sie vorzugsweise über 500  
15 NTs. Die Anzahl der zu analysierenden NSKFs richtet sich  
dabei nach der Gesamtlänge der zu untersuchenden Sequenz,  
der durchschnittlichen Länge der NSKF-Sequenzen und der  
notwendigen Genauigkeit der Sequenzierung. Bei einer durch-  
schnittlichen Länge der ermittelten NSKF-Sequenz von 100  
20 NTs, einer Gesamtlänge der zu untersuchenden Sequenz von 1  
Mb und einer Genauigkeit, die der Rohsequenzermittlung  
entspricht (d.h. jede Stelle soll möglichst nur einmal  
sequenziert werden) benötigt man z.B. die ca. 5-fache Menge  
an Rohsequenzen, d.h. 5 Mb, weil die Verteilung der NSKFs  
25 über die Gesamtsequenz zufällig erfolgt. Insgesamt müssen  
50.000 NSKFs analysiert werden, um mehr als 99% der  
Gesamtstrecke abzudecken.

Anschließend werden die ermittelten NSKF-Sequenzen mit  
30 Hilfe eines kommerziell erhältlichen Programms der  
Vollsequenz zugeordnet und eventuelle Abweichungen  
detektiert. Einem solchen Programm kann z.B. BLAST oder  
FASTA Algorithmus zugrunde liegen ("Introduction to  
computational Biology" 1995 M.S. Waterman Chapman & Hall).

35

Beispiel 4:

Sequenzanalyse mit 2 markierten NTs' und 2 unmarkierten NTs (2NTs' / 2NTs-Methode).

5

In einer anderen Ausführungsform werden für die Analyse der Sequenzen 2 modifizierte NTs' und 2 unmodifizierte NTs eingesetzt.

10 Diese Methode eignet sich besonders zur Analyse der Sequenzvarianten (z.B. SNP- oder Mutationsanalyse) und setzt die Kenntnis einer Referenzsequenz voraus. Dabei wird die Vollsequenz nicht rekonstruiert, sondern die ermittelten Sequenzen werden mit Hilfe eines Programms der Referenzsequenz  
15 zugeordnet und eventuelle Abweichungen registriert. Einem solchen Programm kann z.B. der BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to computational Biology" 1995 M.S. Waterman Chapman & Hall).

20 Diese Ausführungsform beruht auf dem Prinzip, dass eine Abfolge aus 2 Signalen (markierte NT's) genügend Informationen zur Identifizierung einer Sequenz enthalten kann. Die ermittelte Sequenz wird mit der Referenzsequenz verglichen und einer bestimmten Position zugeordnet, z.B.:

25

ACCAAAACACCC - ermittelte Sequenz (dCTP' und dATP' sind markiert)

ATCATCGTTCGAAATATCGATCGCCTGATGCC - Referenzsequenz

30

A-C---C-AAA-A-C-A-C-CC (zugeordnete ermittelte Sequenz)  
ATCATCGTTCGAAATATCGATCGCCTGATGCC (Referenzsequenz)

35 Die unbekannte, zu analysierende Variante der Referenzsequenz wird wie oben beschrieben zur Sequenzierung vorbereitet (NSK wird in NSKFs überführt, diese werden mit PBS ligiert, anschließend mit einem Primer hybridisiert und auf



- 60 -

Reaktionsoberfläche immobilisiert). Auf diese Weise vorbereitete NSKFs werden mit 2NTs'/2NTs-Methode sequenziert. Man erhält NSKF-Sequenzen, wobei jede NSKF-Sequenz eine Abfolge aus 2NTs' darstellt. Um eine eindeutige Zuordnung der 5 ermittelten Sequenz zu einer bekannten Referenzsequenz zu ermöglichen, muß diese Abfolge lang genug sein. Vorzugsweise beträgt die Länge der ermittelten NSKF-Sequenzen mehr als 40 NT's. Da 2 markierte NTs' nur einen Teil der Sequenz darstellen, ist die Gesamtlänge des synthetisierten komplementären Strangs 10 ca. doppelt so lang, wie die Abfolge der detektierten NTs' (bei 40 detektierten NTs' beträgt die Gesamtlänge z.B. durchschnittlich 80 NTs).

Zur Synthese eines komplementären Stranges werden 4 Nukleotide 15 benötigt. Da die mit einem Fluoreszenzfarbstoff markierten NTs' in der vorliegenden Erfindung als Semiterminatoren auftreten, d.h. die Termination ausschließlich bei Verfügbarkeit modifizierter NTs' auftritt, müssen unmodifizierte NTs in einem zusätzlichen Schritt in jedem Zyklus in die Reaktion zugegeben 20 werden. Die genaue Position dieses Schrittes in dem Zyklus kann variieren. Wichtig dabei ist, dass die markierten NTs' und die unmodifizierte NTs getrennt verwendet werden.

Ein Zyklus bei dieser Ausführungsform kann beispielhaft folgen- 25 dermaßen aussehen:

- a) Zugabe einer Lösung mit modifizierten NTs' und Polymerasen auf die Oberfläche mit den bereitgestellten NSKFs
- b) Inkubation der immobilisierten Nukleinsäureketten mit 30 dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind
- c) Waschen
- d) Detektion der Signale von einzelnen, modifizierten und in die den NSKFs komplementären neusynthetisierten Strängen 35 eingebauten NTs'-Molekülen
- e) Entfernung der Markierung und der terminierenden Gruppe bei den eingebauten Nukleotiden
- f) Waschen

- 61 -

- g) Zugabe von 2 unmodifizierten NTs und Polymerasen
- h) Waschen.

5 Diese 2NT's/2NTs-Methode eignet sich beispielsweise für die SNP-Analyse einer genomischen Strecke eines Gens oder für doppelsträngige cDNA-Analyse. Ihr liegen folgende Prinzipien zugrunde:

- 10 1) Die genetische Information in jedem der beiden komplementären DNA-Stränge ist identisch, so dass fehlende Informationen in einem Strang durch die Information aus dem anderen Strang vervollständigt werden können.
- 2) Durch bestimmte Paarkombinationen markierter NTs' kann man  
15 mit nur 2 NTs' die komplette Information aus einer doppelsträngigen DNA erhalten. Zulässige Kombinationen von NT's bei dieser Ausführungsform sind: A'C'; A'G'; C'T'/C'U'; G'T'/G'U'. Bevorzugt wird die Kombination C' und U'.
- 3) Als Grundlage der Analyse dient eine bereits bekannte Referenzsequenz.  
20
- 4) Die NSKFs stammen von beiden Strängen der zu analysierenden NSK und die ermittelten NSKF-Sequenzen decken die gesamte Länge der zu analysierenden Sequenz ab.

25 Am folgenden Beispiel wird erklärt, wie die Information aus einem doppelsträngigen DNA-Fragment mit nur 2 markierten NTs' gewonnen wird und wie die Unterschiede zur ursprünglichen oder nicht mutierten Sequenz (Referenzsequenz / Vergleichsequenz) festgestellt werden können. Sequenzen unter (1) und (2) sind  
30 bis auf eine Stelle identisch (unterstrichen). A' und C' sind markiert.

1) zu prüfende Sequenz:

35 Die zu prüfende Sequenz wird mit 2NT's/2NTs-Methode sequenziert, so dass eine Population an NSKF-Sequenzen (ermittelte NSKF-Sequenzen(n)) entsteht. Diese ermittelten NSKF-Sequenzen enthalten Information von jedem Strang:

- 62 -

5 'A-C---C-AAA-A-C-A-C-CC3' - ermittelte NSKF-Sequenz(i)  
 5 'ATCGTTTCGAAATATCGATCGCCTG3'  
 3 'TAGCAAGCTTTATAGCTAGCGGAC5'  
 5 3' A-CAA-C---A-A-C-A-C---C5' - ermittelte NSKF-Sequenz(i+1)

## 2) Vergleichsequenz:

Zur Analyse ist eine Vergleichsequenz (Referenzsequenz)  
 10 erforderlich:

5 'ATTGTTTCGAAATATCGATCGCCTG3'  
 3 'TAACAAGCTTTATAGCTAGCGGAC5'

## 15 3) Vergleichsequenz mit angepaßten ermittelten NSKF-Sequenzen:

Mit Hilfe eines Programms werden ermittelte NSKF-Sequenzen  
 bestimmten Stellen in der Vergleichsequenz zugeordnet und  
 eventuelle Abweichungen detektiert:

20  
 5 'A-C---C-AAA-A-C-A-C-CC3' - ermittelte NSKF-Sequenz(i)  
 5 'ATTGTTTCGAAATATCGATCGCCTG3'  
 3 'TAACAAGCTTTATAGCTAGCGGAC5'  
 3 ' A-CAA-C---A-A-C-A-C---C5' - ermittelte NSKF-Sequenz(i+1)  
 25     ↑  
       (Einzelnukleotidmutation)

Mit dieser Ausführungsform kann man eine doppelsträngige  
 Nukleinsäure auf SNP oder Mutationen untersuchen. Dabei werden  
 30 die ermittelten NSKF-Sequenzen mit einer Referenzsequenz  
 verglichen. Die Grundregeln des Vergleichs einer Teilsequenz  
 und einer kompletten Sequenz bei der Analyse mit nur 2  
 markierten NTs unterscheiden sich nicht prinzipiell von denen,  
 die bei dem Vergleich der Sequenzen anhand aller 4 markierten  
 35 NTs' gelten. Näheres s. Sequenzvergleich bei Mutationsanalyse  
 und SNP-Analyse mit 4NTs' (Beispiel 3B).

Beispiel 5:

Eine besondere Ausführungsform des Verfahrens stellt die Analyse von Einzelnukleotidpolymorphismen mit 5 sequenzspezifischen Primern dar.

Zusätzlich zu Abschnitt 1 "Abkürzungen und Begriffserläuterungen" werden für dieses Beispiel folgende Begriffe definiert:

10

Primer - Zur Verdeutlichung des erfinderischen Gedankens werden in diesem Beispiel folgende Begriffe unterschieden:

15

a) Unter einem „Primer“ wird vorliegend allgemein eine Population von Primermolekülen mit einheitlicher Struktur verstanden.

b) „mehrere Primer“ o.ä. werden im Text als mehrere Populationen von Primermolekülen verstanden, die unterschiedliche Struktur besitzen.

20

c) Ein „Primer-Molekül“ bedeutet ein einziges Oligonukleotid-Molekül.

d) „Mehrere Primer-Moleküle“ bedeuten mehrere einzelne Oligonukleotid-Moleküle; sie können einheitliche oder unterschiedliche Struktur aufweisen.

25 SNP-Stelle - eine Position in NSK, die auf Vorhandensein oder Abwesenheit von SNP untersucht wird.

Zielsequenz - Teil einer Gesamtsequenz, der durch die Verwendung eines spezifischen Primers in der Sequenzierungsreaktion sequenziert/ bestimmt wird. Eine Gesamtsequenz kann mehrere Zielsequenzen enthalten. Eine Zielsequenz ist genügend lang, um eine Positionierung dieser Zielsequenz innerhalb der Gesamtsequenz mit großer Wahrscheinlichkeit zu gewährleisten. Zielsequenzen können  
35 beispielsweise eine oder mehrere SNP-Stellen enthalten.

Erkennungssequenz - Teil der Zielsequenz, der für die Zuordnung dieser Zielsequenz in der Gesamtsequenz verwendet

wird.

In dieser Ausführungsform zur SNP-Analyse werden mehrere potentielle SNP-Positionen in der Referenzsequenz ausgewählt, 5 die in einer zu analysierenden NSK untersucht werden. Zu diesen Positionen werden entsprechend unterschiedliche, sequenzspezifische Primer bereitgestellt. Diese Primer können einen standardisierten Primersatz zur SNP-Analyse bei einer bestimmten Fragestellung bilden und einheitlich als Kit für 10 die betreffende Analysen eingesetzt werden.

Die Vorbereitung des zu analysierenden Materials (auf SNP zu untersuchende einzel- und doppelsträngige Nukleinsäureketten) hat erfindungsgemäß das Ziel, aus einer oder mehreren langen 15 Nukleinsäureketten (Gesamtsequenz) eine Population an relativ kleinen, zwischen 30 und 2000 NT langen, einzelsträngigen Nukleinsäurekettenfragmenten (NSKFs) zu bilden.

Diese NSKF-Moleküle werden zufällig auf einer planen 20 Oberfläche mit einer Dichte zwischen 10 und 1.000.000 pro 100  $\mu\text{m}^2$ -, vorzugsweise 10 und 100 NSKFs pro 100  $\mu\text{m}^2$ , 100 bis 10.000 pro 100  $\mu\text{m}^2$  oder 10.000 bis 1.000.000 pro 100  $\mu\text{m}^2$  immobilisiert. An die auf der Oberfläche gebundenen NSKFs werden Primer hybridisiert, so dass die Dichte der extensionsfähigen NSKF- 25 Primer-Komplexe ca. 10-100 pro 100  $\mu\text{m}^2$  beträgt. Nach der Hybridisierung werden nicht gebundene Primer entfernt und die Sequenzierungsreaktion gestartet.

Durch eine Auswahl der Zielsequenzen und der 30 sequenzspezifischen Primer werden nur die relevanten Abschnitte der Gesamtsequenz untersucht, was die Menge nicht relevanter Informationen verringert und die Analysezeit verkürzt.

35 Dieser Ausführungsform des Verfahrens zur SNP-Analyse liegen folgende Prinzipien zugrunde:

Es werden Stellen in einer Referenzsequenz ausgewählt, die



- 65 -

in den zu untersuchenden NSKs (Gesamtsequenz) auf Einzelnukleotidpolymorphismen (SNPs) überprüft werden sollen.

- 5 1) Zur Analyse jeder ausgewählten SNP-Stelle werden spezifische Primer bereitgestellt, so dass jede zu untersuchende SNP-Stelle entweder die nächste Position in 3 -Richtung vom Primer einnimmt oder innerhalb von 2 bis 100, vorzugsweise 2 bis 50, idealerweise 2 bis 20  
10 Positionen in 3 -Richtung vom Primer liegt. Die SNP-Stelle liegt somit innerhalb der Zielsequenz, die während der Sequenzierungsreaktion bestimmt wird. Es werden vorzugsweise mehrere SNP-Stellen gleichzeitig analysiert, so dass mehrere spezifische Primer verwendet werden  
15 müssen. Die Primer werden vorzugsweise so ausgewählt, dass sie möglichst einheitliche Annealing-Temperaturen haben, d.h. Unterschiede zwischen Schmelztemperaturen einzelner Primerpopulationen liegen beispielsweise innerhalb eines Bereichs von ca. 4 Grad, besser innerhalb  
20 von 2 Grad, noch besser innerhalb von 1 Grad.
- 2) Von der Gesamtsequenz werden kurze Nukleinsäurekettenfragmente (NSKFs) abgeleitet, wobei diese Fragmente einzelsträngig sind und eine Länge von  
25 20 bis 2000 NT, vorzugsweise 30 bis 500 NT besitzen.
- 3) NSKF-Moleküle werden in einer zufälligen Anordnung auf der Oberfläche immobilisiert.
- 30 4) Nach der Hybridisierung (Annealing) von sequenzspezifischen Primern an die auf der Oberfläche immobilisierten NSKFs wird eine zyklische Sequenzierungsreaktion durchgeführt, wobei für jedes an der Reaktion beteiligte NSKF-Molekül eine Zielsequenz  
35 ermittelt wird. Die Sequenzierungsreaktion läuft an vielen Molekülen gleichzeitig ab.
- 5) Die ermittelten Zielsequenzen enthalten Information über

- 66 -

die Zugehörigkeit zu einem bestimmten Abschnitt in der Gesamtsequenz und über den SNP in diesem Abschnitt bei der zu untersuchenden Probe. Die Länge der Zielsequenzen und somit die Zahl der Zyklen ist so zu wählen, dass eine Identifizierung der Sequenzen gewährleistet werden kann.

In einer vorteilhaften Ausführungsform werden die ermittelten Zielsequenzen mit der Referenzsequenz verglichen und durch Sequenzübereinstimmung zugeordnet. Bei einer genügend langen ermittelten Zielsequenz kann man sie mit großer Wahrscheinlichkeit zu einer bestimmten Position in der Referenzsequenz zuordnen. Beispielsweise kann eine Sequenz aus 10 NTs mehr als  $10^6$  verschiedene Kombinationen bilden und somit mit einer großen Wahrscheinlichkeit in einer NSK von nur 100.000 NT eindeutig identifiziert werden. Nach der Zuordnung der ermittelten Zielsequenz zur bestimmten Position innerhalb der Referenzsequenz werden Unterschiede in den Sequenzen, die SNPs, sichtbar.

Zur Identifizierung der Zielsequenzen wird in einer anderen vorteilhaften Ausführungsform sowohl die bereits bekannte Anzahl der Primer, ihre Zusammensetzung und ein bereits bekannter, an die Primerbindungsstelle anschließender Sequenzabschnitt der Referenzsequenz verwendet. Dabei werden die ermittelten Zielsequenzen nach ihrer Zugehörigkeit zu den Primern analysiert, wobei nur die nah an der Primerbindungsstelle liegenden Sequenzen berücksichtigt werden müssen. Wenn beispielsweise nur 1000 Primer verwendet werden, reichen weniger als 10 NTs der ermittelten Zielsequenzen, um eine Zuordnung zu den entsprechenden Primern zu ermöglichen.

Die zu analysierende Probe enthält meistens mehrere identische Gesamtsequenzmoleküle, z.B. mehrere Kopien von genomischer DNA aus Zellen eines Gewebes oder mehrere identische mRNA-Populationen aus Zellen eines Gewebes.

a) *Wahl der SNP-Stelle*

Mit der erfindungsgemäßen Methode kann man sowohl bekannte SNP-Stellen analysieren als auch neue SNP-Stellen ermitteln. Als  
5 potentielle SNP-Stelle kann jede Position in der NSK auftreten. Die Auswahl richtet sich nach der Fragestellung, z.B. SNP-Analyse in Genen, deren Produkte mit bestimmten Krankheiten assoziiert sind, oder SNP-Analyse in konservierten, kodierenden Abschnitten der Gene, die für Membranrezeptoren kodieren, oder  
10 Überprüfung bekannter SNP-Stellen in regulatorischen Sequenzen von Genen, die für die Zellteilung wichtig sind.

Eine zu analysierende SNP-Stelle liegt innerhalb einer Zielsequenz, die während der Sequenzierungsreaktion bestimmt  
15 wird. Man kann mehrere SNP-Stellen innerhalb einer Zielsequenz ermitteln. Man kann andererseits auch mehrere Zielsequenzen z.B. innerhalb eines Gens wählen. Wichtig dabei ist, dass die Zielsequenzen in genügendem Abstand voneinander in der Gesamtsequenz liegen. Dieser Abstand ist notwendig, damit nur  
20 ein sequenzspezifischer Primer pro NSKF hybridisiert, und er ist von der durchschnittlichen NSKF-Länge abhängig: je kürzer die NSKFs, desto näher aneinander können Zielsequenzen liegen. Die SNP-Stellen können bei angemessener Primer-Wahl an beiden Strängen einer doppelsträngigen Nukleinsäurekette analysiert  
25 werden.

Das Verfahren bietet auch die Möglichkeit, beispielsweise mehrere SNP-Stellen aus vielen Individuen (als Stichprobe einer Population) gleichzeitig zu kontrollieren. Dadurch kann  
30 z.B. das SNP-Profil einer Population untersucht werden.

b) *Primer für die Sequenzierungsreaktion*

Sequenzierungsreaktion an einem einzelnen NSKF-Molekül wird durch ein Primer-Molekül ermöglicht. Ein sequenzspezifischer  
35 Primer ist erfindungsgemäß notwendig, um die Sequenzierungsreaktion jeweils an einer bestimmten / spezifischen Zielsequenz innerhalb der Gesamtsequenz durchführen zu können. Der für die Analyse einer SNP-Stelle,

- 68 -

bzw. einer Zielsequenz einzusetzende sequenzspezifische Primer stellt eine Population von Primer-Molekülen mit identischer Struktur dar. Für die Analyse mehrerer, unterschiedlicher Zielsequenzen sind mehrere unterschiedliche Primer-  
5 Populationen notwendig.

Durch die Verwendung sequenzspezifischer Primer werden nur die relevanten Sequenzabschnitte, die Zielsequenzen, analysiert. Im erfindungsgemäßen Verfahren wird die zu sequenzierende  
10 Länge der Sequenzen möglichst niedrig gehalten, damit die Geschwindigkeit der Analyse steigt.

Ein sequenzspezifischer Primer bindet an eine für ihn spezifische Primerbindungsstelle in der zu analysierenden  
15 Sequenz, PBS. Die Zusammensetzung und die Länge der Primer werden für jede potentielle SNP-Stelle, bzw. Zielsequenz, optimiert. Beispiele für Optimierungsschritte sind in Rychlik et al. NAR 1990 v.18 S.6409 dargestellt. Bei der Primerwahl bzw. bei der Wahl der PBS (Primerbindungsstelle) sind folgende  
20 Aspekte besonders zu berücksichtigen:

1) Die zu analysierende SNP-Stelle sollte entweder gleich nach dem 3'-Ende des Primers oder innerhalb der nächsten 2 bis 50 NTs, vorzugsweise 2 bis 20 NTs liegen .

25

2) Die Positionierung (die Wahl der Sequenzlänge und der Zusammensetzung) der PBS zu SNP-Stelle sollte so erfolgen, dass die verschiedenen PBS-Sequenzen und die korrespondierenden Primer-Sequenzen möglichst ähnliche  
30 „Annealing-Temperaturen“ besitzen, um bei möglichst einheitlichen Hybridisierungsbedingungen zu binden. Das kann beispielsweise durch Veränderung der PBS-Position im Bezug auf die jeweilige, zu analysierende SNP-Stelle oder durch die Veränderung der Primersequenzlänge erfolgen (Rychlik et  
35 al. NAR 1990 v.18 S.6409).

3) Der minimale Abstand zwischen Primern, die an denselben Strang in der Gesamtsequenz binden, sollte die

durchschnittliche NSKF-Länge nicht unterschreiten.

Es können Primer für beide Stränge einer Doppelstrang-NSK verwendet werden. Damit lassen sich beispielsweise nah  
5 aneinander liegende SNP-Stellen erfassen, oder man kann eine Kontrolle einer SNP-Stelle in beiden Strängen vornehmen.

Vorzugsweise beträgt die Länge des Primers zwischen 6 und 100 NTs, optimalerweise zwischen 10-30 oder 30-40 oder 40-50. Für  
10 verschiedene SNP-Stellen, bzw. Zielsequenzen können Primer mit unterschiedlicher Länge eingesetzt werden.

Für die SNP-Analyse mit sequenzspezifischen Primern werden Primer erfindungsgemäß in einer Hybridisierungslösung an die  
15 auf der Reaktionsoberfläche immobilisierten NSKFs hybridisiert (Annealing-Reaktion).

*c) Immobilisierung von NSKFs*

In dieser Ausführungsform werden erfindungsgemäß die NSKF-  
20 Primer-Komplexe ausschließlich über die NSKFs an die Oberfläche gebunden (direkte Bindung von NSKFs an die Oberfläche), wobei die bereitgestellten NSKF-Moleküle an die plane Oberfläche in zufälliger Anordnung gebunden werden.

25 Die Immobilisierung der NSKFs erfolgt vorzugsweise an einem der beiden Ketten-Enden (s.o.). Die Immobilisierung kann auch durch eine unspezifische Bindung, wie z.B. durch Austrocknung der NSKFs enthaltenden Probe auf der planen Oberfläche erreicht werden. Die Dichte der Immobilisation kann zwischen  
30 10 und 100, 100 und 10.000, 10.000 und 1.000.000 NSKFs pro 100  $\mu\text{m}^2$  liegen.

*d) Hybridisierung*

Die gebundenen NSKFs und die Primer werden unter stringenten  
35 Hybridisierungsbedingungen inkubiert, die eine möglichst selektive Anbindung (Annealing) der Primer an die entsprechenden Primerbindungsstellen der NSKFs erlauben. Optimale Hybridisierungsbedingungen hängen von der genauen



- 70 -

Struktur der Primerbindungsstellen und der jeweiligen Primerstrukturen ab und lassen sich beispielsweise nach Rychlik et al. NAR 1990 v.18 S.6409 berechnen.

5 Die Primer stellen vorzugsweise ein Primergemisch dar. Die Konzentrationen einzelner sequenzspezifischer Primer (Einzelkonzentrationen von Primer-Populationen) liegen beispielsweise zwischen 10pmol/l und 1mmol/l, vorzugsweise zwischen 0.1µmol/l und 10µmol/l. Die Gesamtkonzentration von  
10 Primern im Primergemisch liegt vorzugsweise zwischen 1nmol/l und 10mmol/l. Das Verhältnis zwischen einzelnen Primer-Populationen kann variieren. Primer können in deutlichem Überschuss über die immobilisierten NSKFs zugegeben werden, so dass die Hybridisierungszeit gering ist.

15

Die für die Detektion notwendige Dichte von extensionsfähigen NSKF-Primer-Komplexen beträgt ca. 10 bis 100 pro 100 µm<sup>2</sup>. Sie kann vor, während oder nach der Hybridisierung der Primer erreicht werden.

20

Bei einer bekannten NSKF-Konzentration können in einer Ausführungsform die Immobilisierungsbedingungen so gewählt werden, dass die NSKFs in einer Dichte von ca. 10 bis 1000 Moleküle pro 100µm<sup>2</sup> gebunden werden. NSKFs bestimmen somit die  
25 Dichte der NSKF-Primer-Komplexe.

In einer anderen Ausführungsform kann die Dichte der immobilisierten NSKFs wesentlich höher als 1000 NSKFs pro 100µm<sup>2</sup> liegen, z.B. 1.000.000 pro 100µm<sup>2</sup>. Die für die optische  
30 Detektion notwendige Dichte der NSKF-Primer-Komplexe wird während der Primer-Hybridisierung erreicht. Dabei sind die Hybridisierungsbedingungen (z.B. Temperatur, Zeit, Puffer) so zu wählen, dass die Primer nur an einen Teil der immobilisierten NSKFs binden, s. Beispiel 6.

35

Bei unbekannter NSKF-Konzentration und entsprechend unbekannter Immobilisationsdichte kann die Hybridisierung (Annealing) von Primern an die NSKFs zu einer höheren als

optimale Dichte von NSKF-Primer-Komplexen führen.

Aus diesem Grund wird in einer vorteilhaften Ausführungsform ein Teil der NSKFs enthaltenden Probe für die Ermittlung der optimalen Dichte verwendet. Dieser Teil wird auf einer Reaktionsoberfläche immobilisiert, die Primer werden an die NSKFs hybridisiert und die entstandenen NSKF-Primer-Komplexe werden durch den Einbau von Fluoreszenzfarbstoff tragenden NT\*s (z.B. Cy3-dCTP, Amersham Pharmacia Biotech) markiert. Aus der ermittelten Dichte lässt sich einerseits die eventuell notwendige Verdünnung oder Konzentrierung der ursprünglichen Probe für den endgültigen Sequenzierungsansatz errechnen (Die Hybridisierungsbedingungen werden beibehalten). Andererseits können daraus notwendige Veränderungen in den Hybridisierungsbedingungen errechnet werden, beispielsweise eine Verkürzung der Hybridisierungszeit, wobei die NSKF-Immobilisierungsdichte konstant bleibt.

Das Mengen-Verhältnis zwischen Primerpopulationen kann unterschiedlich oder gleich groß sein. Durch eine höhere Primerkonzentrationen können gewisse, beispielsweise seltenere Sequenzen mit größerer Wahrscheinlichkeit in einem bestimmten Zeitraum gebunden werden.

Der große Vorteil der beschriebenen Verfahrensanordnung gegenüber einer Verfahrensanordnung mit auf einer Oberfläche immobilisierten sequenzspezifischen Primern und einer anschließenden Hybridisierung von Proben an diese Primer ist die deutliche Verkürzung der Zeit für die Hybridisierung (Annealing) zwischen den sequenzspezifischen Primern und den zu analysierenden Proben auf der Reaktionsoberfläche.

Beispiel 6:

Die Vorbereitung und die Durchführung der Sequenzierungsreaktion

5

Die Vorbereitung der Gel-Oberfläche:

Das Polyacrylamid-Gel für die Analyse von Reaktionen mit einzelnen Molekülen wird nach allgemeinen Regeln der Gel-  
10 Vorbereitung für elektrophoretische Auftrennung erstellt („Electrophoresis“ A.T. Andrews, Oxford science publications 1995).

Die Polymerisationsreaktion kann z.B. durch UV-Licht oder  
15 durch Radikalbildner durchgeführt werden. In diesem Beispiel wird Ammoniumpersulfat (APS) und TEMED (Tetramethylethyldiamin) zur Radikalreaktion verwendet, z.B. TEMED 0.01% v/v und APS 0.04% w/v. Die Komponentenzusammensetzung kann breit variieren, die  
20 Konzentrationen einzelner Komponenten liegen in folgenden Bereichen (errechnet für die gebrauchsfertige wässrige AA-bisAA-Lösung):

Acrylamid-Monomer (AA) von 3 bis 30%, idealerweise zwischen 10  
25 und 20% Bis-Acrylamid (bis-AA) im Verhältnis zum Acrylamid-Monomer 1:10 bis 1:50, vorzugsweise 1:20.

Zur Herstellung werden 2 saubere Glasplatten verwendet (mit Aceton und danach Wasser gewaschen). Eine Glasplatte (P1) wird  
30 vorzugsweise mit einem wasserabweisenden Reagenz vorbehandelt, z.B. Repel-silan, Dimethyldichlorsilane-Lösung, Amersham Pharmacia-Biotech. P2 dient als fester Träger für das Gel und kann mit gelbindenden Reagenzien z.B. Bind-silan, Methacryloxypropyltrimethoxysilane, Amersham Pharmacia-  
35 Biotech, vorbehandelt werden, so dass es zu einer kovalenten Bindung zwischen dem Gel und der Glasoberfläche kommt. Die P2-Vorbehandlung mit gelbindenden Reagentien ist dann sinnvoll, wenn mehrere Reaktionen mit immobilisierten Molekülen

- 73 -

durchgeführt werden müssen. Bei einer geringeren Anzahl an Reaktionen ist eine solche Vorbehandlung nicht notwendig. In diesen Fällen reicht, für P2 eine saubere Glas-Oberfläche aus, so dass das Gel allein durch adhäsive Kräfte an der Glasoberfläche haften bleibt.

Die fertige Polymerisationslösung (AA/bisAA-Lösung mit Radikalbildnern) wird zwischen P1 und P2 gegossen, so dass eine Schicht mit der Dicke von ca. 5 bis 30  $\mu\text{m}$  resultiert. Die Dicke des Gels kann z.B. durch Abstandhalter kontrolliert werden. Nach Erhärtung wird P1 entfernt. Das Gel bleibt auf P2 haften. Es wird mit entionisiertem Wasser gewaschen.

Das Gel kann direkt weiter verwendet werden oder in verschiedenen Fertigungsstadien getrocknet und gelagert werden. Vor einer Reaktion mit markierten Molekülen wird das Gel normalerweise einige Minuten in der Reaktions-Pufferlösung aufgequollen und erst dann für die Reaktion eingesetzt.

Auf eine so vorbereiteten Gel-Oberfläche werden NSKFs durch das Austrocknen immobilisiert.

Beispielsweise wurde eine Lösung (ca.  $1\mu\text{l}$ ) einer Plasmid-DNA (mit Hind III linearisierte, durch Hitze in einzelsträngige Form überführte pMOS-Blue-Plasmid-DNA ca. 3400 NT lang, Konzentration  $0.1\mu\text{g}/\mu\text{l}$ ) auf ca.  $10\text{mm}^2$  der Gel-Oberfläche aufgetropft und bei  $90^\circ\text{C}$  zum Trocknen gebracht. Die errechnete Dichte der immobilisierten Plasmid-Moleküle betrug ca. 1000 pro  $1\mu\text{m}^2$ .

Als Primer wurde das Oligonukleotid 5'-AGTGAATTCGAGCTCGGTAC-3' verwendet. Die Primerbindungsstelle (nachfolgend fettgedruckt) zusammen mit der für die Analyse relevanten Verlängerung hat folgende Sequenz:

5'-ATCCCCGGGTACCGAGCTCGAATTCACT-3'

Eine Flow-Cell (Mikroflüssigkeitskanal, MFK), Fig.6, mit der

- 74 -

Reaktionsoberfläche als Deckel wurde zusammengebaut. Ein solcher MFK erlaubt einen schnellen Flüssigkeitsaustausch unter der Geloberfläche.

5 Als Vorversuch wurde der Primer (errechnete  $T_m$  45,3° C, 0.1  $\mu\text{mol/l}$  in 50mmol/l Tris-HCl pH 8,7) bei 45° C für 10 Minuten mit der Plasmid-DNA auf der Oberfläche hybridisiert (Annealing). Nach einem Waschschrift wurde die Dichte der Plasmid-Primer-Komplexe kontrolliert. Die Kontrolle erfolgte  
10 durch den Einbau von dCTP-Cy3 (Amersham Pharmacia Biotech) unter Verwendung von Klenow-Fragment (2Units pro 50  $\mu\text{l}$  in 20mmol/l Tris-HCl-Puffer, pH 8,5, mit 5mmol/l  $\text{MgCl}_2$ , 15 Minuten bei 30° C). Dabei wird nur ein einzelnes dCMP-Cy3 in den wachsenden Strang eingebaut.

15

Als Detektionsapparatur diente Axioplan 2e (Zeiss) mit der CCD-Kamera AxioCam (Zeiss), Fig. 8.

Die Signaldichte der einzelnen, eingebauten dCMP-Cy3-Moleküle  
20 entspricht der Dichte der extensionsfähigen Plasmid-Primer-Komplexe. Unter den genannten Bedingungen betrug die Dichte der Plasmid-Primer-Komplexe durchschnittlich ca. 15 pro 100  $\mu\text{m}^2$  und lag damit in der gewünschten Größenordnung (Fig. 8a-c).

Auf einer zweiten, in gleicher Weise vorbereiteten Oberfläche  
25 (mit Hind III linearisierte, durch Hitze in einzelsträngige Form überführte pMOS-Blue-Plasmid-DNA ca. 3400 NT lang, Konzentration 0.1  $\mu\text{g}/\mu\text{l}$  mit hybridisierten Primern) wird eine zyklische Sequenzierungsreaktion durchgeführt. Dabei werden  
30 (dCTP\*) (s. Beispiel 2) als reversible Terminatoren verwendet. Die Detektionsapparatur ist dieselbe wie im Vorversuch. Die für die zyklische Sequenzierungsreaktion verwendeten Lösungen setzen sich wie folgt zusammen:

35 a) Reaktionslösung für die Einbaureaktion: 20 mmol/l Tris-HCl-Puffer, pH 8.5, 5mmol/l  $\text{MgCl}_2$ , 10% Glycerin, Klenow-Fragment (Amersham Pharmacia-Biotech) 2U pro 50  $\mu\text{l}$ , dUTP\* bzw. dCTP\*, oder dATP und dGTP je 10  $\mu\text{mol/l}$ .

- 75 -

b) Waschlösung: 20 mmol/l Tris-HCl pH 8.5, 0.01% Na-Azid

c) Reaktionslösung für die Abspaltungsreaktion: 20 mmol/l  
5 Tris-HCl, pH 8.5, 50mmol/l Mercaptoethanol.

Die Einbaureaktionen mit markierten NT\*s wurden bei 30°C für 15 Minuten durchgeführt.

- 10 Im ersten Zyklus der Sequenzierungsreaktion wurde eine Reaktionslösung mit dCTP\* zugegeben. Nach einem Waschschrift wurde ein Detektionsschritt durchgeführt, wobei Einzelmolekül-Signale mit den zugeordneten x,y-Koordinaten auf der Oberfläche registriert wurden (insgesamt ca. 11.200 Signale).
- 15 Danach wurde die Markierung von den eingebauten NT\*s abgespalten (Raumtemperatur, 10 Minuten) und die Oberfläche gewaschen.

Im zweiten Zyklus wurde eine Reaktionslösung mit dUTP\*  
20 zugegeben und 15 Minuten lang bei 30° C inkubiert. Nach einem anschließenden Waschschrift wurden die Einzelmolekül-Signalen auf der Oberfläche detektiert (insgesamt ca. 200 Signale). Dies entspricht dem Hintergrundsignal, das durch eine unspezifische Bindung der NT\*s an die Oberfläche entsteht. Die  
25 Markierung von den NT\*s wurde abgespalten (Raumtemperatur, 10 Minuten) und die Oberfläche mit der Waschlösung gewaschen.

Im dritten Zyklus wurde eine Reaktionslösung mit dATP und dGTP zugegeben und 15 Minuten lang bei 30° C inkubiert. Anschließend  
30 wurde die Oberfläche gewaschen.

Die Zyklen 1 bis 3 wurden drei mal wiederholt, wobei insgesamt ca. 9900 CCU-Zielsequenzen ermittelt wurden. Diese Sequenzen können eindeutig dem Primer zugeordnet werden.



Legenden zu den Figuren 1 bis 9

## Legende zu Fig. 1

## 5 Schematische Darstellung der Sequenzierung einer langen Nukleinsäurekette

Der Sequenzierung und der Rekonstruktion von langen Nukleinsäuresequenzen (NSKs) liegt das Shotgun-Prinzip 10 zugrunde. Die Sequenz eines langen DNA-Stücks wird dabei durch die Sequenzierung kleiner Fragmente (NSKFs) und eine nachfolgende Rekonstruktion ermittelt.

1) Ausgangsmaterial - die zu analysierende lange 15 Nukleinsäuresequenz, Gesamtsequenz

2) Fragmente von 50-1000 bp - die im Fragmentierungsschritt aus der Gesamtsequenz erzeugten NSKFs

20 3) Fragmente mit jeweils einem Primer - NSKF-Primer-Komplexe

4) Immobilisierte Fragmente - an die plane Oberfläche gebundene NSKF-Primer-Komplexe, in dieser Ausführungsform erfolgt die Bindung am 3'-Ende der NSKFs

25

5) Zugabe einer Lösung mit Polymerasen und NT\*s - der erste Schritt in einem Zyklus der Sequenzierungsreaktion

6) Waschschrift - nach dem Einbauschrift wird die Oberfläche 30 gewaschen

7) Detektion - die Signale von einzelnen eingebauten NT\*s werden detektiert

35 8) Entfernung der Markierung und der zur Termination führenden Gruppe - zur Fortsetzung der Sequenzierungsreaktion werden die Markierung und das sterische Hindernis entfernt

**Legende zur Fig. 2****Beispiele für allgemeine Struktur von NSKF-Primer-Komplexen**

5

**Fig. 2a** In dieser Ausführungsform wird eine einheitliche Primerbindungsstelle (PBS) an das 3'-Ende der NSKFs angekoppelt und an diese PBS bindet ein einheitlicher Primer.

10 **Fig. 2b** In dieser Ausführungsform werden natürlich vorkommende Primerbindungsstellen verwendet, daher werden Gemische spezifischer Primer verwendet

1) Primer

2) Primerbindungsstelle

15 3) NSKF

**Legende zur Fig. 3**

Ein Beispiel für die Ankopplung einer einheitlichen  
20 Primerbindungsstelle (PBS), die eine funktionelle Gruppe zur Bindung an die Oberfläche trägt.

In diesem Fall wird ein doppelsträngiger Oligonukleotidkomplex (3a), der beispielsweise eine Modifikation an beiden Strängen  
25 hat (3b), an die doppelsträngigen NSKFs liegt (3c). Nach Denaturierung entstehen einzelsträngige NSKFs mit einheitlicher PBS (3d).

30 **Legende zur Fig. 4**

Ein anderes Beispiel für die Erzeugung einer einheitlichen Primerbindungsstelle (PBS).

35 In diesem Fall werden NTs an das 3'-Ende der einzelsträngigen NSKFs angekoppelt (ein so genanntes „Tailing“). Durch Verwendung eines einheitlichen NT entsteht eine einheitliche PBS.

**Legende zur Fig. 5****5 Beispiel für die Bindung von NSKFs an eine gelartige Reaktionsoberfläche.**

Auf einer festen Unterlage (1) haftet eine Gelschicht (2), z.B. ein Polyacrylamidgel (Fig. 5a), oder haften viele  
10 Gelkügelchen (5), z.B. Agarosekügelchen (Fig. 5b). An die Oberfläche des Gels sind NSKFs (4) gebunden. Die NSKFs tragen eine funktionelle Gruppe, z.B. Biotin, und sind an das Gel über Streptavidin oder Avidin (3) gebunden.

15

**Legende zur Fig. 6****Beispiel für eine Durchflussvorrichtung**

20 Eine gelartige Reaktionsoberfläche (1) ist auf einer für das Anregungs- und Fluoreszenzlicht durchlässigen festen Unterlage (2) befestigt. Sie bilden zusammen den Deckel der Flow-Cell. Die Flüssigkeiten in der Flow-Cell können kontrolliert ausgetauscht werden, wobei die Flow-Cell zusammen mit  
25 Vorratsbehälter (3), Pumpe (4) und Ventil (5) eine Durchflussvorrichtung bilden. Auf der Reaktionsoberfläche sind NSKF-Primer-Komplexe gebunden (hier nicht abgebildet). Die Signale der eingebauten NT\*s werden mit der Detektionsapparatur (6) detektiert.

30

**Legende zur Fig. 7**

**Strukturen von 2'-Deoxynukleosidtriphosphaten, die im Verfahren eingesetzt werden können.**

35

**Fig. 7a - Schematische Darstellung der NT-Struktur, bei der die spaltbare Gruppe und die sterisch anspruchsvolle, zur Termination führende Gruppe Teile des Linkers bilden. Der**

- 79 -

Linker ist die Verbindung zwischen Nukleobase und Fluoreszenzfarbstoff.

A,B,C,D,E - Linker, A - der Linkerrest nach der Spaltung, B - spaltbare Gruppe, D - sterisch anspruchsvolle, zur Termination 5 führende Gruppe, F - Fluoreszenzfarbstoff.

Fig. 7b - Schematische Darstellung der NT-Struktur, wobei die spaltbare Gruppe ein Teil des Linkers ist und der Fluoreszenzfarbstoff gleichzeitig die sterisch anspruchsvolle, 10 zur Termination führende Gruppe darstellt.

A,B,C - Linker, A - der Linkerrest nach der Spaltung, B - spaltbare Gruppe, D - sterisch anspruchsvolle, zur Termination führende Gruppe, F - Fluoreszenzfarbstoff.

15 Fig. 7c - Schematische Darstellung der Struktur von eingebauten NT\*s nach dem Abspaltungsschritt. Dargestellt sind zwei NT\*s mit dem verbliebenen Linkerrest (A).

Fig. 7d - Schematische Darstellung der NT-Struktur, wobei die 20 spaltbare Gruppe, die gleichzeitig die sterisch anspruchsvolle, zur Termination führende Gruppe ist, einen Teil des Linkers darstellt.

A,B,C,D - Linker, A - der Linkerrest nach der Spaltung, B - spaltbare Gruppe, D - sterisch anspruchsvolle, zur Termination 25 führende Gruppe, F - Fluoreszenzfarbstoff.

Fig. 7e - Darstellung von bevorzugten NT-Strukturen, bei denen der Linker an die 5-Position im Pyrimidinring angekoppelt ist.

30 Fig. 7f - Darstellung anderer bevorzugter NT-Strukturen, bei denen der Linker an die 5-Position im Pyrimidinring angekoppelt ist.

Fig. 7g - Darstellung von bevorzugten NT-Strukturen, bei denen 35 der Linker an die 7-Position im Purinring angekoppelt ist.

Fig. 7h,i,j - Beispiele für die Ankopplung von Farbstoffen an den Linker

**Fig. 7k**

Strukturen von weiteren 2'-Deoxynukleosidtriphosphaten, die im Verfahren eingesetzt werden können. Der Linker ist an die 5-Position des Pyrimidinrings gekoppelt.

Die Substituenten  $R_{1,2,3,4}$  sind wählbar und können unabhängig voneinander auftreten.

10 Die Z-Gruppe stellt in einer Ausführungsform (7k-1) die Verbindung zwischen dem Linker und der Base dar. Sie ist wählbar und kann eine Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminogruppe sein.

15 Die E-Gruppe stellt in einer Ausführungsform (7k-1) einen internen Teil des Linkers dar. In einer anderen Ausführungsform (7k-2) stellt sie die Verbindung zwischen dem Linker und der Base dar. Diese Gruppe ist wählbar und kann eine unverzweigte Alkyl- oder Alkenylkette mit einer Zahl von 20 Kohlenstoffatomen, vorzugsweise zwischen 1 und 5, sein.

Die E-Gruppe kann aber auch eine Alkyl- oder Alkenylkette mit einer internen Amid- Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung sein.

25 Die C-Gruppe ist in diesem Beispiel eine chemisch spaltbare Gruppe. In den Ausführungsformen (7k-1,2) stellt sie einen internen Teil des Linkers dar. In einer anderen Ausführungsform (7k-3) stellt sie die Verbindung zwischen dem Linker und der Base dar. Diese Gruppe ist wählbar und kann 30 eine Ester-, Thioester- und Disulfidverbindung sein.

Die Y-Gruppe stellt einen internen Teil des Linkers dar, der die Verbindung zwischen der spaltbaren Gruppe (C) und dem Fluoreszenzfarbstoff (F) herstellt. Diese Gruppe ist wählbar 35 und kann eine verzweigte oder unverzweigte Alkyl- oder Alkenylkette oder auch eine substituierte oder unsubstituierte Arylgruppe sein. Eine weitere mögliche Alternative ist eine Alkyl- oder Alkenylkette mit einer internen Amid- Carbalcoxy-

(Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung.

Die X-Gruppe ist die Verbindung zwischen dem Fluoreszenzfarbstoff und dem Linker, wobei diese Verbindung  
5 sowohl vom Linker, als auch vom Fluoreszenzfarbstoff (F) abgeleitet werden kann. Sie ist wählbar und kann eine Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminogruppe sein.

10 Fig. 7L

Strukturen von weiteren 2'-Deoxynukleosidtriphosphaten, die im Verfahren eingesetzt werden können. Der Linker ist an die 7-Position des Purinrings gekoppelt.

15 Die Substituenten  $R_{1,2,3,4}$  sind wählbar und können unabhängig voneinander auftreten.

Die Z-Gruppe stellt in einer Ausführungsform (7L-1) die Verbindung zwischen dem Linker und der Base dar. Sie ist  
20 wählbar und kann eine Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminogruppe sein.

Die E-Gruppe stellt in einer Ausführungsform (7L-1) einen internen Teil des Linkers dar. In einer anderen  
25 Ausführungsform (7L-2) stellt sie die Verbindung zwischen dem Linker und der Base dar. Diese Gruppe ist wählbar und kann eine unverzweigte Alkyl- oder Alkenylkette mit einer Zahl von Kohlenstoffatomen, vorzugsweise zwischen 1 und 5, sein.

Die E-Gruppe kann aber auch eine Alkyl- oder Alkenylkette mit  
30 einer internen Amid- Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung sein.

Die C-Gruppe ist in diesem Beispiel eine chemisch spaltbare Gruppe. In den Ausführungsformen (7L-1,2) stellt sie einen  
35 internen Teil des Linkers dar. In einer anderen Ausführungsform (7L-3) stellt sie die Verbindung zwischen dem Linker und der Base dar. Diese Gruppe ist wählbar und kann eine Ester-, Thioester- und Disulfidverbindung sein.



Die Y-Gruppe stellt einen internen Teil des Linkers dar, der die Verbindung zwischen der spaltbaren Gruppe (C) und dem Fluoreszenzfarbstoff (F) herstellt. Diese Gruppe ist wählbar und kann eine verzweigte oder unverzweigte Alkyl- oder Alkenylkette oder auch eine substituierte oder unsubstituierte Arylgruppe sein. Eine weitere mögliche Alternative ist eine Alkyl- oder Alkenylkette mit einer internen Amid- Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung.

10

Die X-Gruppe ist die Verbindung zwischen dem Fluoreszenzfarbstoff und dem Linker, wobei diese Verbindung sowohl vom Linker, als auch vom Fluoreszenzfarbstoff (F) abgeleitet werden kann. Sie ist wählbar und kann eine Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminogruppe sein.

**Fig. 7m**

Weitere Beispiele für Strukturen von

20 2'-Deoxynukleosidtriphosphaten, die im Verfahren eingesetzt werden können. Der Linker ist an die 5-Position des Pyrimidinrings gekoppelt.

Die Substituenten  $R_{1,2,3,4}$  sind wählbar und können unabhängig voneinander auftreten.

Die Y-Gruppe stellt einen internen Teil des Linkers dar, der die Verbindung zwischen der spaltbaren Gruppe (C) und dem Fluoreszenzfarbstoff (F) herstellt. Diese Gruppe ist wählbar und kann eine verzweigte oder unverzweigte Alkyl- oder Alkenylkette oder auch eine substituierte oder unsubstituierte Arylgruppe sein. Eine weitere mögliche Alternative ist eine Alkyl- oder Alkenylkette mit einer internen Amid- Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminobindung.

35

Die X-Gruppe ist die Verbindung zwischen dem Fluoreszenzfarbstoff und dem Linker, wobei diese Verbindung sowohl vom Linker, als auch vom Fluoreszenzfarbstoff (F)

- 83 -

abgeleitet werden kann. Sie ist wählbar und kann eine Amid-, Carbalcoxy- (Ester), Sulfoxy-, Ether-, Thioether- oder Aminogruppe sein.

5

Legende zur Fig. 8

#### Beispiel für ein Detektionssystem

10 Dargestellt ist ein Weitfeld-Optik-Detektionssystem. Nach dem Einbau von markierten NT\*s wird die Oberfläche (7) abgescannt, wobei die Fluoreszenzsignale von einzelnen, an die NTs gekoppelten Farbstoffmolekülen detektiert werden.

15 Fig. 8a Schematische Darstellung eines Abschnittes der Reaktionsoberfläche (grau), der abgescannt wird. Die Kreise entsprechen jeweils der Aufnahme eines 2D-Bildes und repräsentieren die Flächen, von denen die Fluoreszenzsignale detektiert werden. Dabei werden pro Aufnahme mehrere Signale  
20 (beispielsweise 100 bis 10.000) von einzelnen Molekülen gleichzeitig registriert. Die Reaktionsoberfläche wird in jedem Zyklus abgescannt, wobei während des Scannvorganges mehrere Aufnahmen von unterschiedlichen Stellen der Oberfläche gemacht werden. Dabei können bis zu mehreren Millionen Signale  
25 von eingebauten NT\*s aufgenommen werden. Die hohe Parallelität ist die Grundlage für die Geschwindigkeit des Verfahrens.

Fig. 8b Eine Aufnahme (ein 2D-Bild) mit Signalen von einzelnen, eingebauten NT\*s. Zur Versuchsbeschreibung siehe  
30 Beispiel 6.

Fig. 8c Ausschnitt aus Abbildung 8b. Der Ausschnitt zeigt Signale von vier eingebauten NT\*s. Jedes Signal besitzt charakteristische Eigenschaften der Einzelmolekülsignale (s.  
35 Beschreibung) und kann auf grund dieser identifiziert werden (vorzugsweise mit Hilfe eines Computer-Programms). Jedem der identifizierten Signale werden die entsprechenden X,Y-Koordinaten zugeordnet.

**Legende zur Fig. 9****5 Beispiel einer vorteilhaften Anordnung von  
Reaktionsoberflächen.**

Der Durchsatz wird durch Verwendung von zwei getrennten Flow-Cells (Mikroflüssigkeitskanäle, MFK) erhöht. Während in der  
10 einen Flow-Cell biochemische und chemische Reaktionen ablaufen, wird in der anderen die Detektion durchgeführt. Anschließend tauschen die Flow-Cells ihre Positionen.

Patentansprüche:

1. Verfahren zur parallelen Sequenzanalyse von Nukleinsäuresequenzen (Nukleinsäureketten, NSKs), bei dem man

Fragmente (NSKFs) einzelsträngiger NSKs mit einer Länge von etwa 50 bis 1000 Nukleotiden erzeugt, die überlappende Teilsequenzen einer Gesamtsequenz darstellen können, man

die NSKFs unter Verwendung eines einheitlichen oder mehrerer unterschiedlichen Primer in Form von NSKF-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

eine zyklische Aufbaureaktion des komplementären Stranges der NSKFs unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

- a) zu den auf der Oberfläche gebundenen NSKF-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs') enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs' jeweils an den NTs' befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs' durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs' strukturell an der Base so modifiziert sind, dass die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT' in denselben Strang einzubauen, wobei der Fluoreszenzfarbstoff abspaltbar ist und die strukturelle Modifikation ein abspaltbarer sterisch anspruchsvoller Ligand ist, man
- b) die in Stufe a) erhaltene stationäre Phase unter

Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT' verlängert werden, man

- c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs' geeignet sind, man
- d) die einzelnen, in komplementäre Stränge eingebauten NTs' durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- e) zur Erzeugung unmarkierter (NTs oder) NSKFs die Fluoreszenzfarbstoffe und die sterisch anspruchsvollen Liganden von den am komplementären Strang angefügten NTs' abspaltet, man
- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Liganden geeignet sind, man

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

wobei man die relative Position einzelner NSKF-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser NSKFs durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus nur jeweils ein markiertes NT' einsetzt.

3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus jeweils zwei unterschiedlich markierte NTs<sup>\*</sup> einsetzt.
4. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus jeweils vier unterschiedlich markierte NTs<sup>\*</sup> einsetzt.
5. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die NSKs Varianten einer bekannten Referenzsequenz sind und man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in den Zyklen abwechselnd jeweils zwei unterschiedlich markierte NTs<sup>\*</sup> und zwei unmarkierte NTs einsetzt und man die Gesamtsequenzen durch Vergleich mit der Referenzsequenz ermittelt.
6. Verfahren nach den Ansprüchen 1 bis 5, dadurch gekennzeichnet, dass man in die NSKFs jeweils eine Primerbindungsstelle (PBS) einführt, wobei man bei doppelsträngigen NSKs an beiden komplementären Einzelsträngen jeweils eine PBS einführt und wobei die Primerbindungsstellen für alle NSKFs jeweils gleiche oder verschiedene Sequenzen aufweisen.
7. Verfahren nach den Ansprüchen 1 bis 6, dadurch gekennzeichnet, dass man die NSKFs mit Primern in einer Lösung unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der NSKFs geeignet sind, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen, und man die gebildeten NSKF-Primer-Komplexe anschließend auf der Reaktionsoberfläche bindet.
8. Verfahren nach den Ansprüchen 1 bis 6, dadurch gekennzeichnet, dass man die NSKFs zunächst auf der Reaktions-



oberfläche immobilisiert und erst anschließend mit Primern unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der NSKFs geeignet sind, wobei NSKF-Primer-Komplexe gebildet werden, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen.

9. Verfahren nach den Ansprüchen 1 bis 6, dadurch gekennzeichnet, dass man die Primer zunächst auf der Reaktionsoberfläche immobilisiert und erst anschließend mit NSKFs unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der NSKFs geeignet sind, wodurch NSKFs an die Oberfläche gebunden und NSKF-Primer-Komplexe gebildet werden, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen.
10. Verfahren nach den Ansprüchen 1 bis 9, dadurch gekennzeichnet, dass die Dichte der extensionsfähigen NSKF-Primer-Komplexe zwischen etwa 10 und 100 pro  $100\mu\text{m}^2$  liegt.
11. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die Nukleinsäuresequenzen (NSKs) Sequenzabschnitte einer Gesamtsequenz sind und die Primer sequenzspezifische Primer sind, wobei man

einzelsträngige NSKFs mit einer Länge von etwa 30 bis 1000 Nukleotiden bereitstellt, die überlappenden Teilsequenzen der Gesamtsequenz entsprechen, man

die NSKF-Moleküle direkt an einer planen Oberfläche in einer zufälligen Anordnung bindet, man

mit einer oder mehreren sequenzspezifischen Primerpopulationen eine Hybridisierung (Annealing) an die immobilisierten NSKFs durchführt, wobei die Dichte der einzelnen extensionsfähigen NSKF-Primer-Komplexe zwischen 10 und 100 pro  $100\mu\text{m}^2$  liegt, man

eine zyklische Aufbaureaktion der zu NSKFs komplementären Stränge durchführt, indem man

- a) zu den gebundenen NSKF-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs') enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs' jeweils an den NTs' befindlichen Fluoreszenzfarbstoffe so gewählt sind, daß sich die verwendeten NTs' durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, und die NTs' strukturell so modifiziert sind, dass die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT' in denselben Strang einzubauen, wobei die zur Termination führende Gruppe mit dem Fluoreszenzfarbstoff abspaltbar ist, man
- b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT' verlängert werden, man
- c) die in Stufe b) erhaltenen stationären Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs' geeignet sind, man
- d) die einzelnen, in komplementäre Stränge eingebauten NT'-Moleküle durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man

- e) die zur Termination führenden Gruppen mit den Fluoreszenzfarbstoffen von den am komplementären Strang angefügten NTs zur Erzeugung unmarkierter (NTs oder) NSKFs abspaltet, man
- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der zu Termination führenden Gruppen mit den Fluoreszenzfarbstoffen geeignet sind, man

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

wobei man die relative Position einzelner NSF-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser NSFs durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt.

- 12. Verfahren nach Anspruch 11, dadurch gekennzeichnet, dass man im Detektionsschritt (d) folgende Detektionsarten einsetzt: Weitfeld-Epifluoreszenzmikroskopie, Laser-Scanning-Fluoreszenzmikroskopie, TIRF-Mikroskopie.
- 13. Verfahren nach Anspruch 11 oder 12, dadurch gekennzeichnet, dass die Konzentration einzelner sequenzspezifischer Primer während der Hybridisierung (Annealing) zwischen 10pmol/l und 1µmol/l liegt.
- 14. Verfahren nach Anspruch 11 oder 12, dadurch gekennzeichnet, dass die Konzentration einzelner sequenzspezifischer Primer während der Hybridisierung (Annealing) zwischen 1µmol/l und 1mmol/l liegt.
- 15. Verfahren nach den Ansprüchen 11 bis 14, dadurch gekennzeichnet, dass es ein Verfahren zur SNP-Analyse ist und man einen sequenzspezifischen Primer zur Identifizierung jeder SNP-Stelle in der Gesamtsequenz

verwendet.

16. Verfahren nach Anspruch 15, dadurch gekennzeichnet, dass die Zahl der parallel zu analysierenden SNP-Stellen größer als 2 ist und man für jede SNP-Stelle einen sequenzspezifischen Primer verwendet.
17. Verfahren nach den Ansprüchen 1 bis 16, dadurch gekennzeichnet, dass der zur strukturellen Modifikation verwendete sterisch anspruchsvolle Ligand der zur Markierung verwendete Fluoreszenzfarbstoff ist.
18. Verfahren nach den Ansprüchen 1 bis 17, dadurch gekennzeichnet, dass die Reaktionsoberfläche aus der Gruppe bestehend aus Silicon, Glas, Keramik, Kunststoffen, Gelen ausgewählt ist.
19. Verfahren nach Anspruch 18, dadurch gekennzeichnet, dass die Kunststoffe Polycarbonate oder Polystyrole oder Derivate derselben sind.
20. Verfahren nach Anspruch 18, dadurch gekennzeichnet, dass die Gele Agarose- oder Polyacrylamidgele oder Derivate derselben sind.
21. Verfahren nach Anspruch 20, dadurch gekennzeichnet, dass die Gele 1 bis 2 % Agarose-Gele oder 10 bis 15 % Polyacrylamid-Gele sind.
22. Verfahren nach den Ansprüchen 1 bis 21, dadurch gekennzeichnet, dass die Polymerase eine DNA-Polymerase ohne 3'-5'-Endonukleaseaktivität ist.
23. Verfahren nach den Ansprüchen 1 bis 21, dadurch gekennzeichnet, dass die Polymerase aus der Gruppe bestehend aus thermolabilen viralen, bakteriellen, eukaryontischen DNA-Polymerasen und thermostabilen bakteriellen Polymerasen ausgewählt ist.

24. Verfahren nach Anspruch 23, dadurch gekennzeichnet, dass die DNA-Polymerase Sequenase Version 2, Klenow-Fragment der DNA-Polymerase I aus E.coli ohne 3'-5' exonukleatische Aktivität, Polymerase Beta aus Säugetieren, Taq-Polymerase oder ProHA-DNA-Polymerase ist.
25. Verfahren nach den Ansprüchen 1 bis 24, dadurch gekennzeichnet, dass die NTs' alpha-Phosphorothioat-NTs' sind und die Polymersase das Klenow-Fragment von E.coli-Polymerase I oder T4-DNA-Polymerase ist.
26. Verfahren nach den Ansprüchen 1 bis 25, dadurch gekennzeichnet, dass die Fluoreszenzfarbstoffe aus der Gruppe bestehend aus Cyanin-Farbstoffen, Rhodamine, Xanthene und deren Derivaten ausgewählt sind.
27. Träger zur Durchführung des Verfahrens nach den Ansprüchen 11 bis 26, dadurch gekennzeichnet, dass auf seiner Oberfläche die NSKFs in einer zufälligen Anordnung immobilisiert sind, wobei die Dichte der immobilisierten NSKF-Moleküle zwischen 10 und 100 pro  $100\mu\text{m}^2$  liegt.
28. Träger zur Durchführung des Verfahrens nach den Ansprüchen 11 bis 26, dadurch gekennzeichnet, dass auf seiner Oberfläche die NSKFs in einer zufälligen Anordnung immobilisiert sind, wobei die Dichte der immobilisierten NSKF-Moleküle zwischen 100 und 1.000.000 pro  $100\mu\text{m}^2$  liegt.
29. Kit zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 26, dadurch gekennzeichnet, dass es eine Reaktionsoberfläche (einen festen Träger), zur Durchführung des Verfahrens erforderliche Reaktionslösungen, ein oder mehrere Polymerasen, und Nukleotide (NTs) enthält, von denen ein bis vier mit Fluoreszenzfarbstoffen markiert sind, wobei die markierten NTs ferner strukturell so modifiziert sind (NT' bzw. NTs'), dass die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in

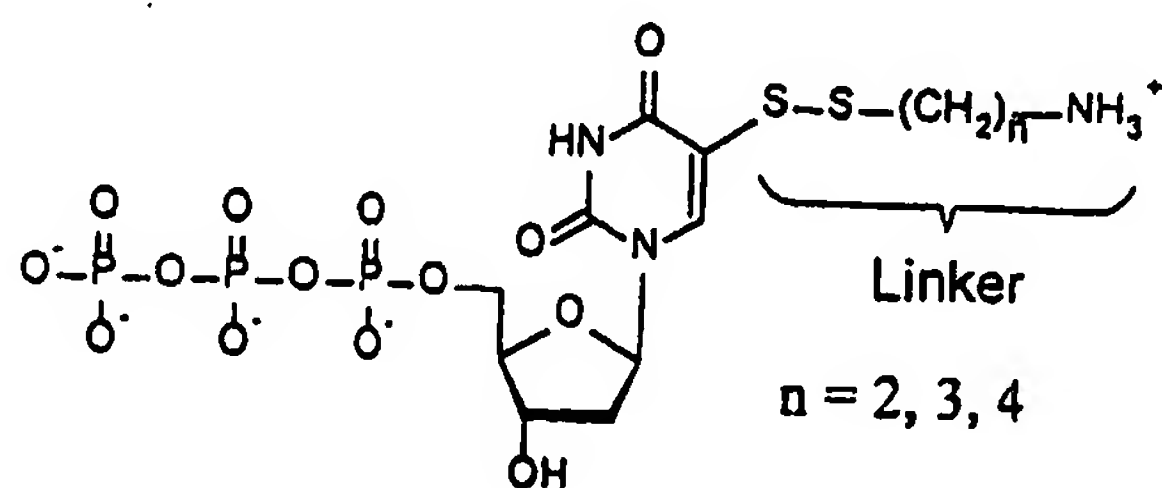
der Lage ist, ein weiteres NT' in denselben Strang einzubauen, wobei der Fluoreszenzfarbstoff abspaltbar ist und die strukturelle Modifikation ein abspaltbarer sterisch anspruchsvoller Ligand ist.

30. Kit nach Anspruch 29, dadurch gekennzeichnet, dass der zur strukturellen Modifikation verwendete sterisch anspruchsvolle Ligand der zur Markierung verwendete Fluoreszenzfarbstoff ist.
31. Kit nach Anspruch 29 oder 30, dadurch gekennzeichnet, dass es ferner Bestandteile enthält:
  - a) zur Erzeugung von Einzelsträngen aus Doppelsträngen erforderliche Reagenzien,
  - b) Nukleinsäuremoleküle, die als PBS in die NSKFs eingeführt werden,
  - c) Oligonukleotid-Primer,
  - d) zur Abspaltung der Fluoreszenzfarbstoffe und sterisch anspruchsvollen Liganden erforderliche Reagenzien, und/oder
  - e) Waschlösungen.
32. Kit nach den Ansprüchen 29 bis 31, dadurch gekennzeichnet, dass die Reaktionsoberfläche aus der Gruppe bestehend aus Silicon, Glas, Keramik, Kunststoffen, Gelen ausgewählt ist.
33. Kit nach Anspruch 32, dadurch gekennzeichnet, dass die Kunststoffe Polycarbonate oder Polystyrole oder Derivate derselben sind.
34. Kit nach Anspruch 32, dadurch gekennzeichnet, dass die Gele Agarose- oder Polyacrylamidgele oder Derivate derselben sind.
35. Kit nach Anspruch 34, dadurch gekennzeichnet, dass die Gele 1 bis 2% Agarose-Gele oder 10 bis 15% Polyacrylamid-Gele sind.

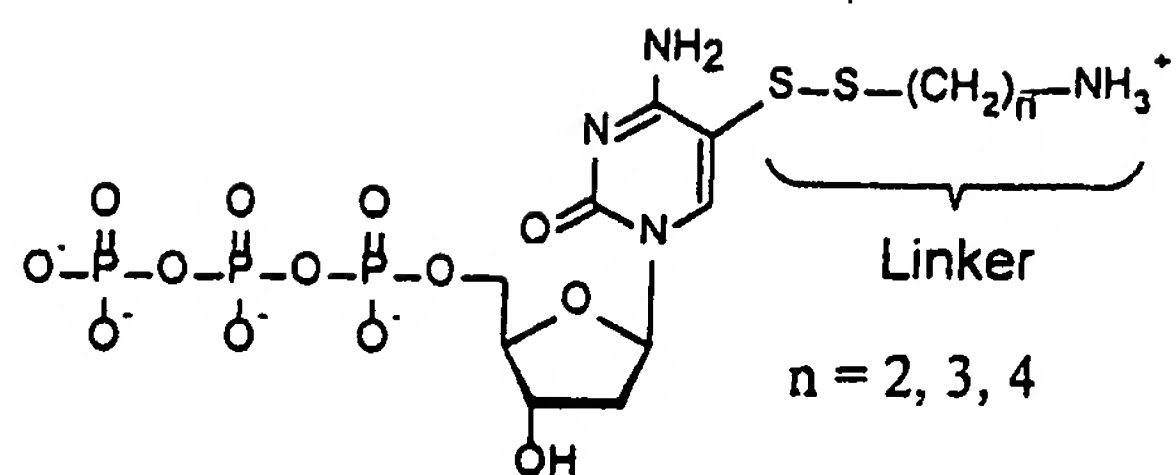


36. Kit nach den Ansprüchen 29 bis 35, dadurch gekennzeichnet, daß die Reaktionsoberfläche ein Träger nach den Ansprüchen 27 und 28 ist.
37. Kit nach den Ansprüchen 29 bis 36, dadurch gekennzeichnet, dass die DNA-Polymerase eine DNA-Polymerase ohne 3'-5'-Endonukleaseaktivität ist.
38. Kit nach Anspruch 37, dadurch gekennzeichnet, dass die Polymerase aus der Gruppe bestehend aus thermolabilen viralen, bakteriellen, eukaryontischen DNA-Polymerasen und thermostabilen bakteriellen Polymerasen ausgewählt ist.
39. Kit nach Anspruch 38, dadurch gekennzeichnet, dass die DNA-Polymerase Sequenase Version 2, Klenow-Fragment der DNA-Polymerase I aus E.coli ohne 3'-5' exonukleatische Aktivität, Polymerase Beta aus Säugetieren, Taq-Polymerase oder ProHa-DNA-Polymerase ist.
40. Kit nach den Ansprüchen 29 bis 39, dadurch gekennzeichnet, dass die NTs' alpha-Phosphorothioat-NTs' sind und die Polymersase das Klenow-Fragment von E.coli-Polymerase I oder T4-DNA-Polymerase ist.
41. Kit nach den Ansprüchen 29 bis 40, dadurch gekennzeichnet, dass die Fluoreszenzfarbstoffe aus der Gruppe bestehend aus Cyanin-Farbstoffen, Rhodamine, Xanthene und deren Derivaten ausgewählt sind.
42. Nukleotid der Formel.

7e-1)

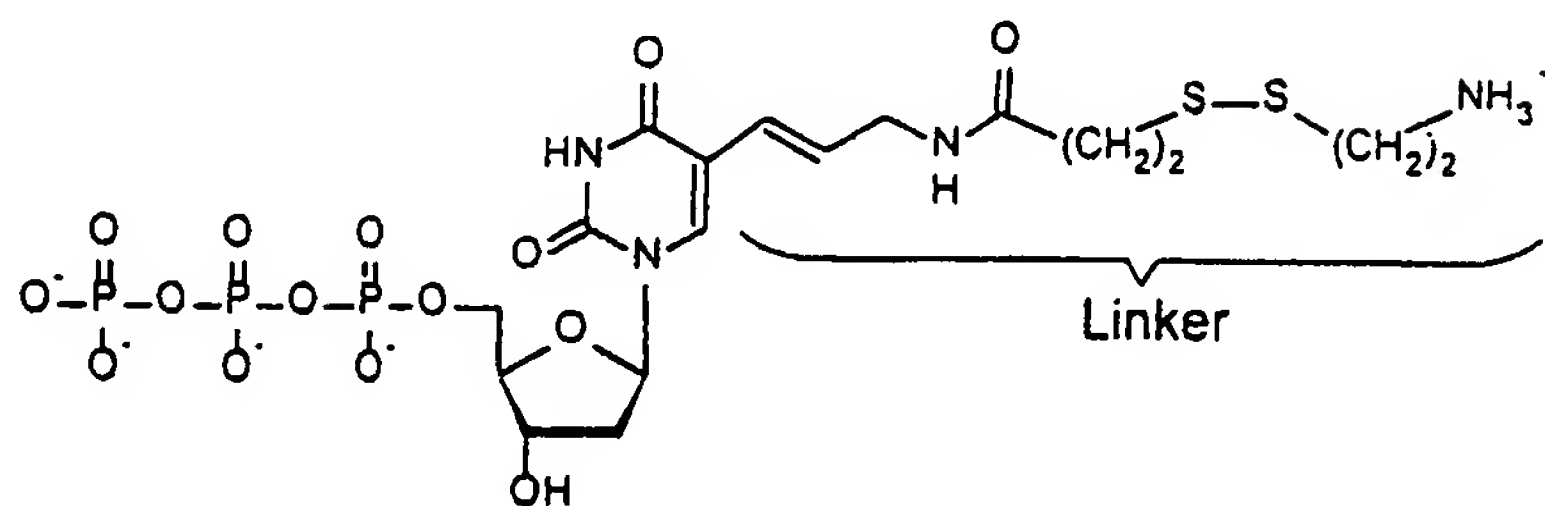


7e-2)

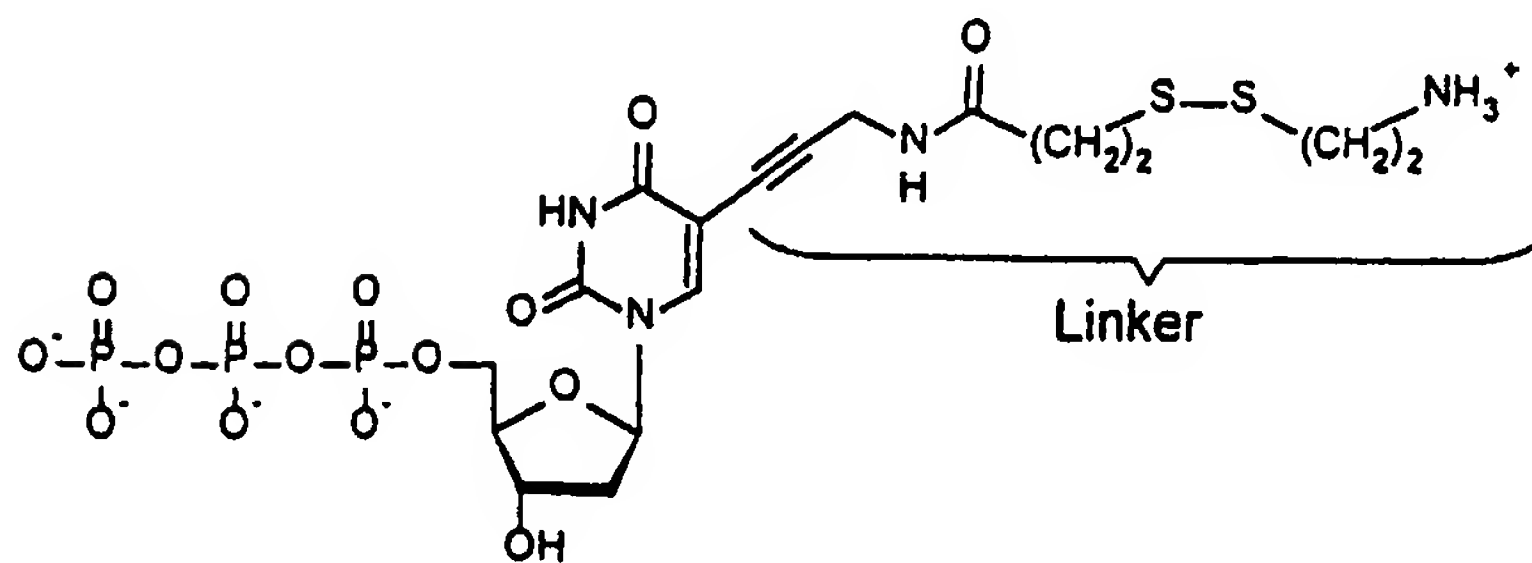


43. Nukleotid der Formel

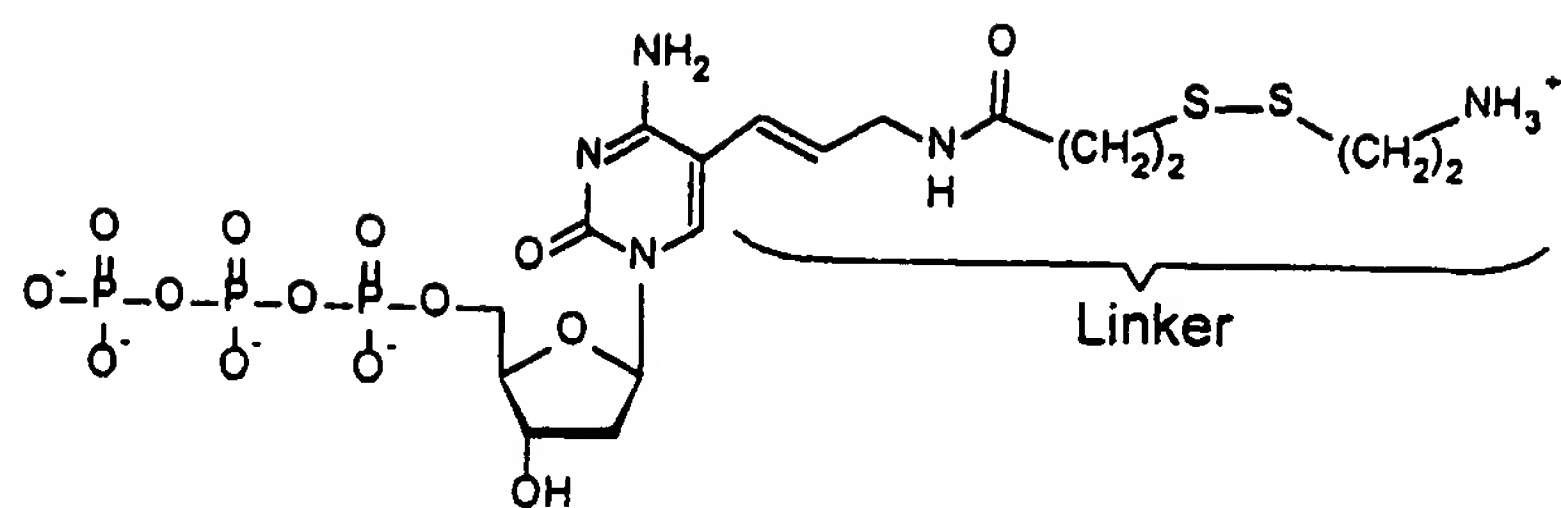
7f-1)



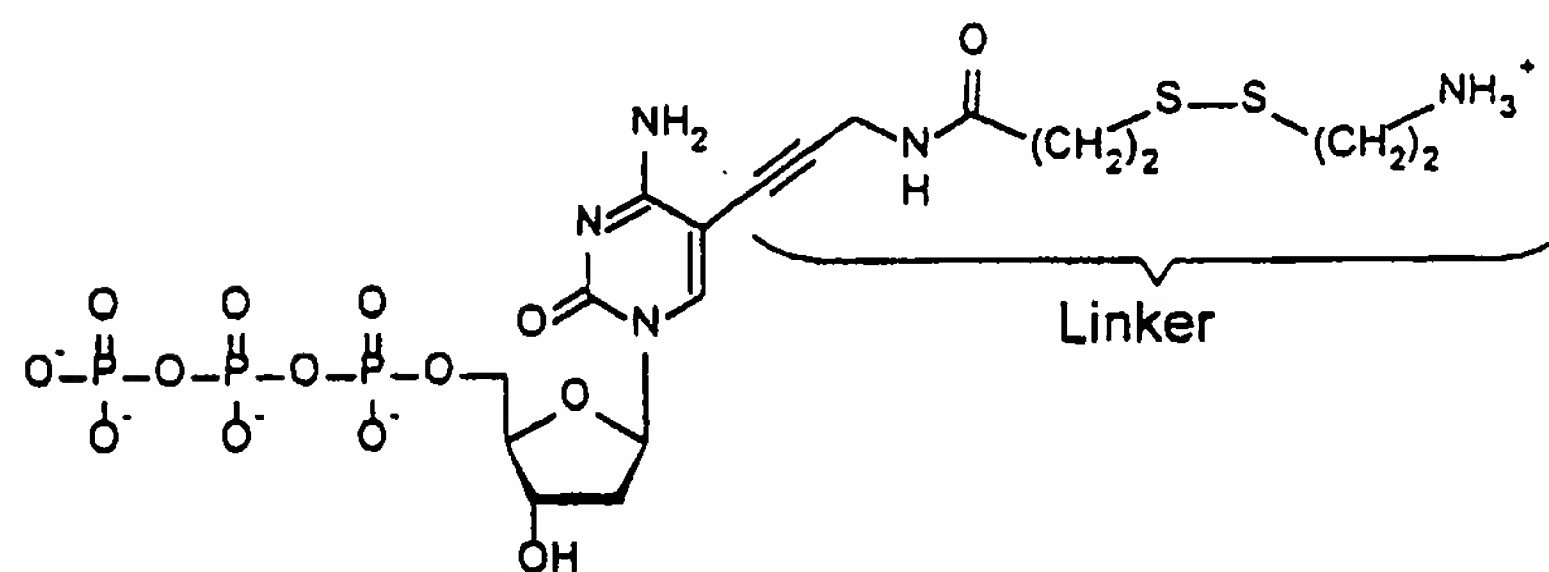
7f-2)



7f-3)

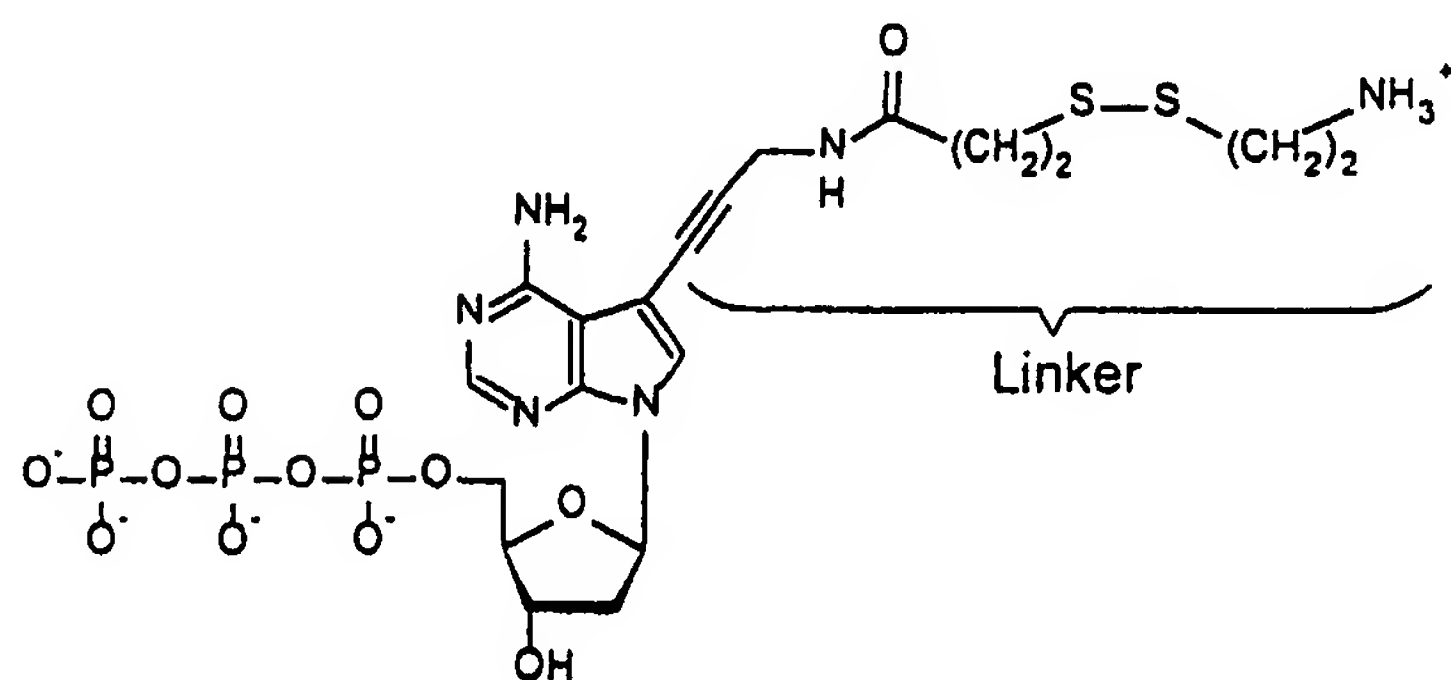


7f-4)

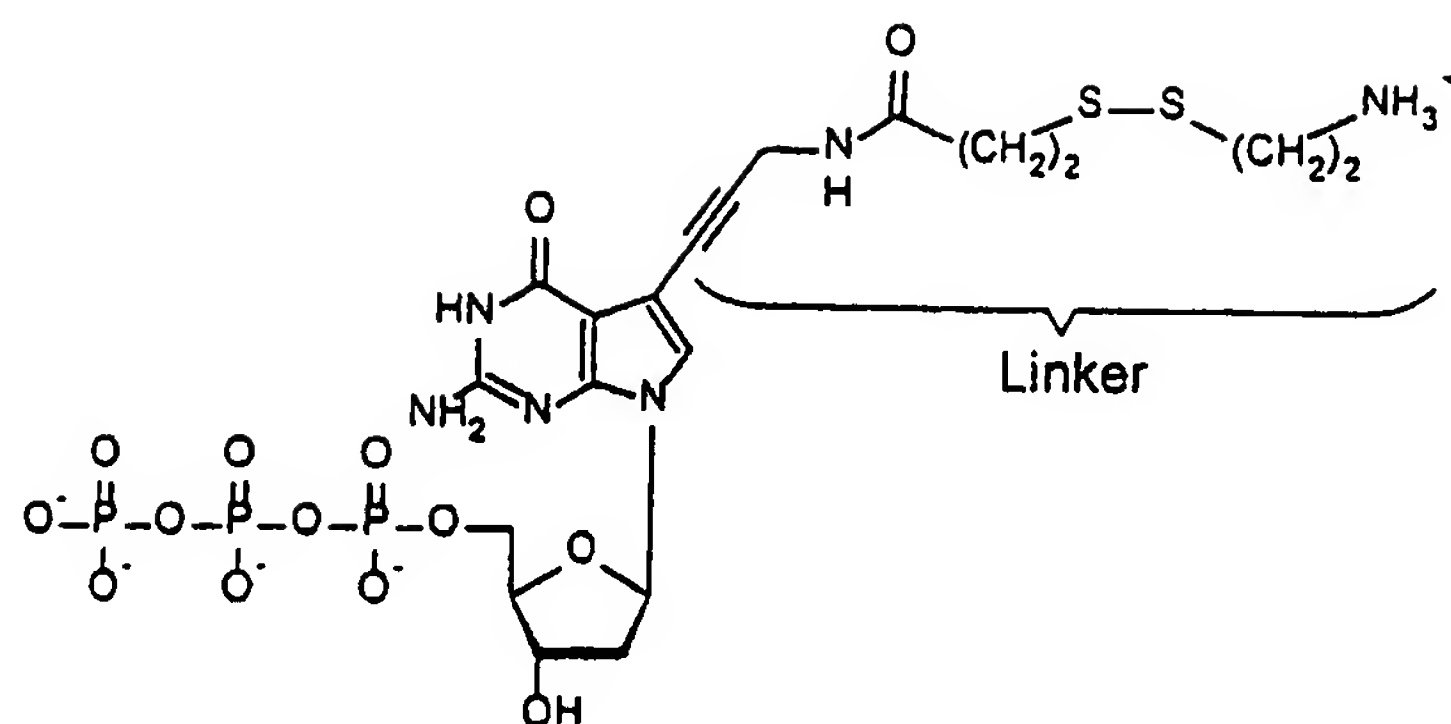


44. Nukleotid der Formel

7g-1)



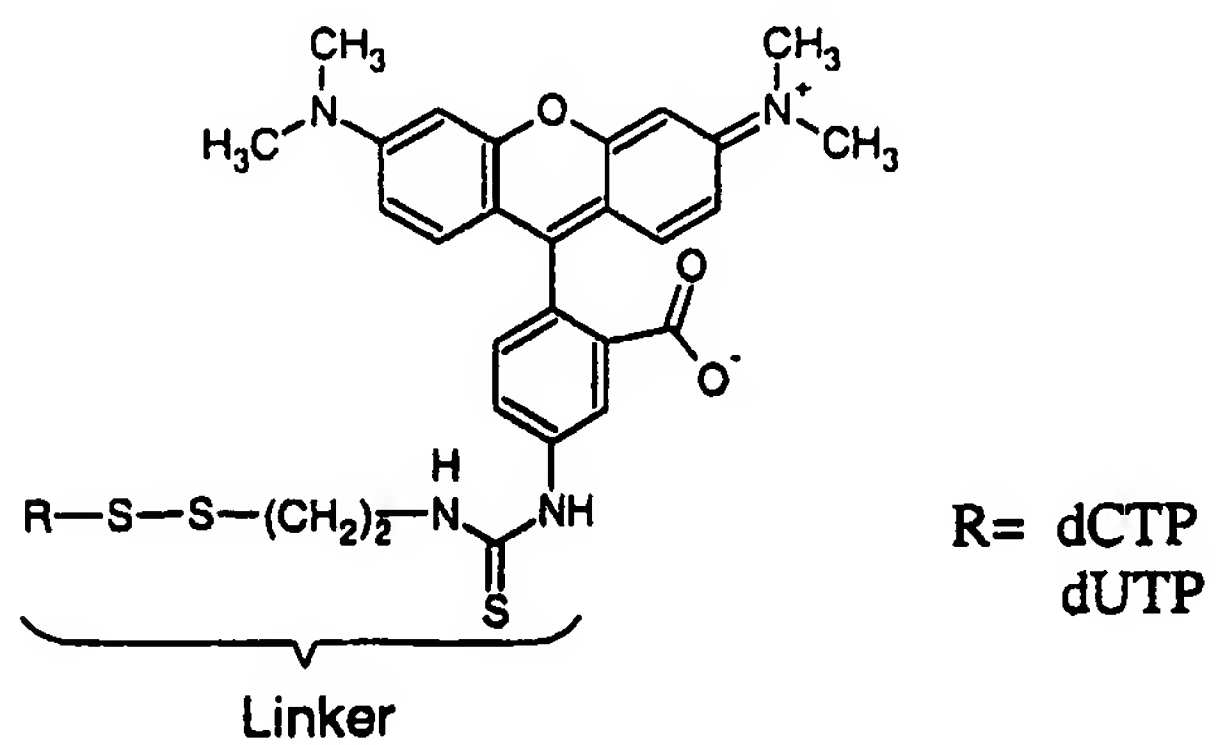
7g-2)



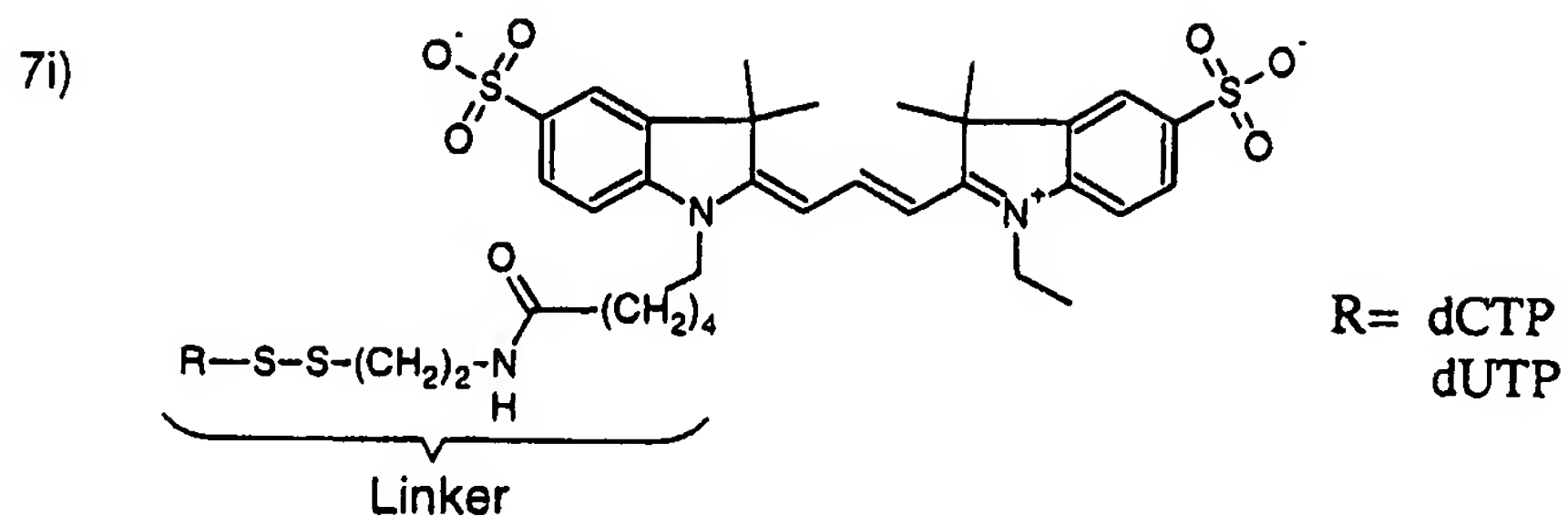
45. Nukleotid nach Anspruch 42 bis 44, dadurch gekennzeichnet, dass an die terminale Aminogruppe ein Fluoreszenzfarbstoff angeheftet ist.

46. Nukleotid der Formel

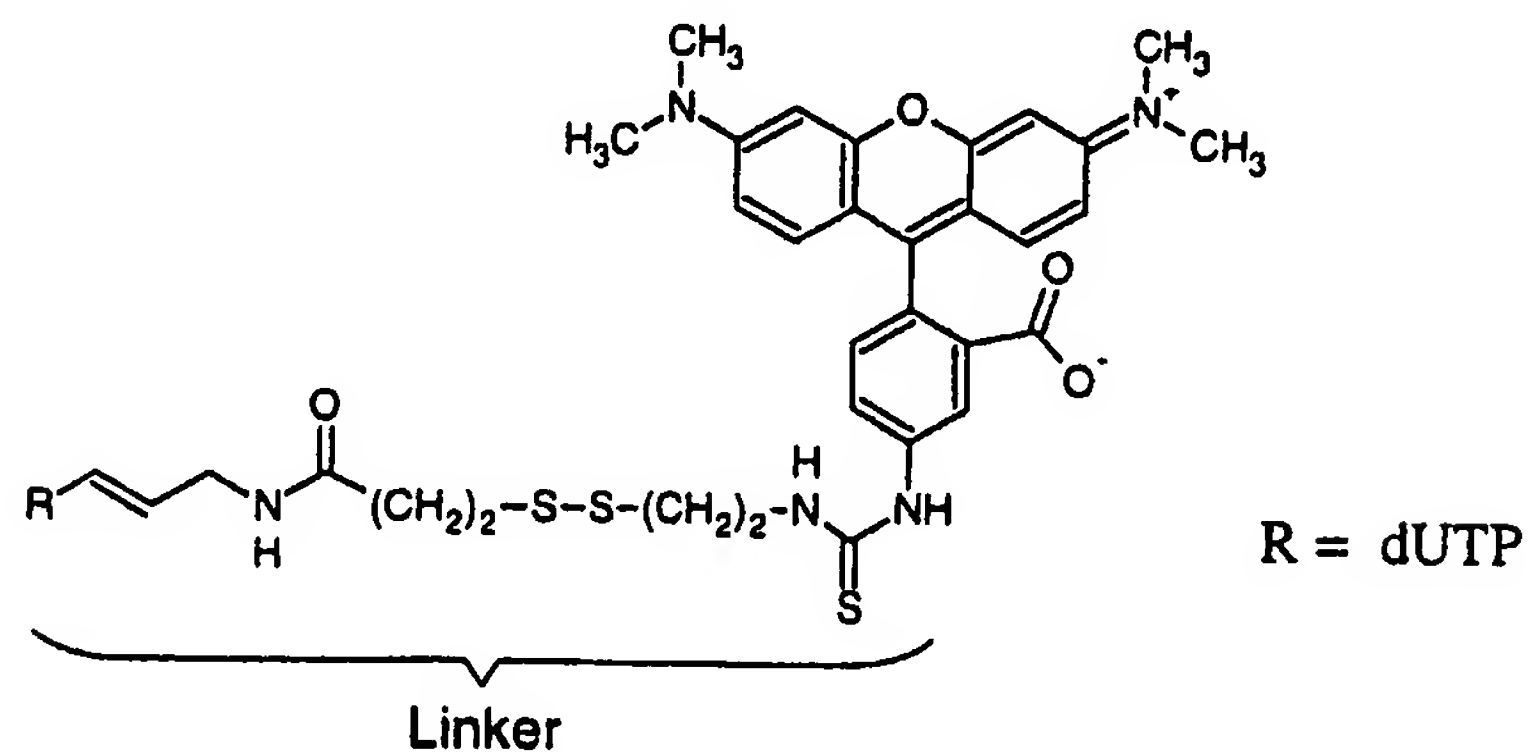
7h)



## 47. Nukleotid der Formel



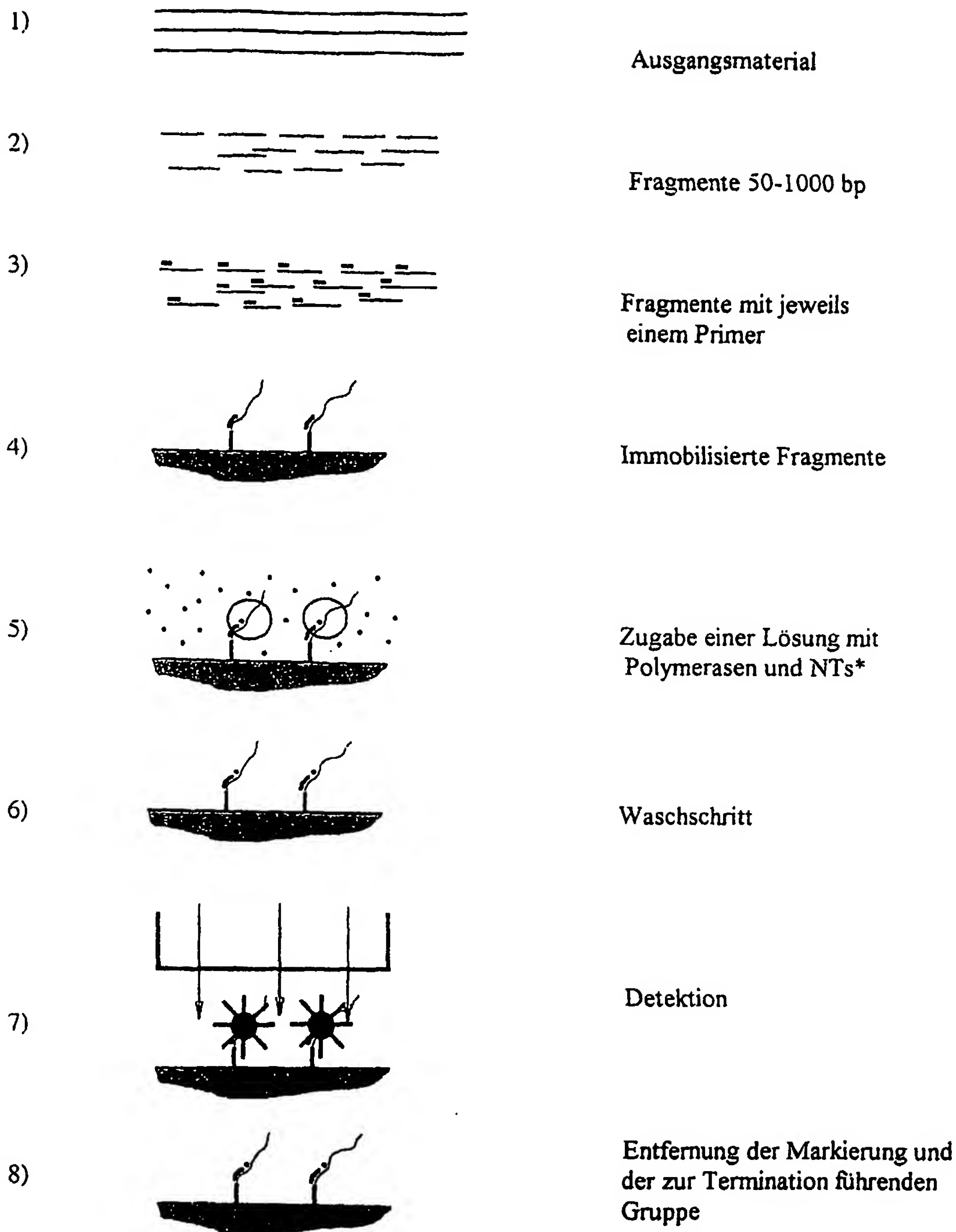
## 48. Nukleotid der Formel

7j) dNTP-SS-TRITC (L 14)

49. Verwendung eines Nukleotids nach den Ansprüchen 42 bis 48 als NT' in einem Verfahren nach den Ansprüchen 1 bis 26.

50. Verwendung eines Nukleotids entsprechend Fig. 7k, 7L, 7m, als NT' in einem Verfahren nach den Ansprüchen 1 bis 26.

Fig. 1



**Legende:**

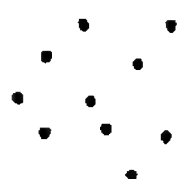
feste Oberfläche

Polymerase

NT\*

Signal von NT\*

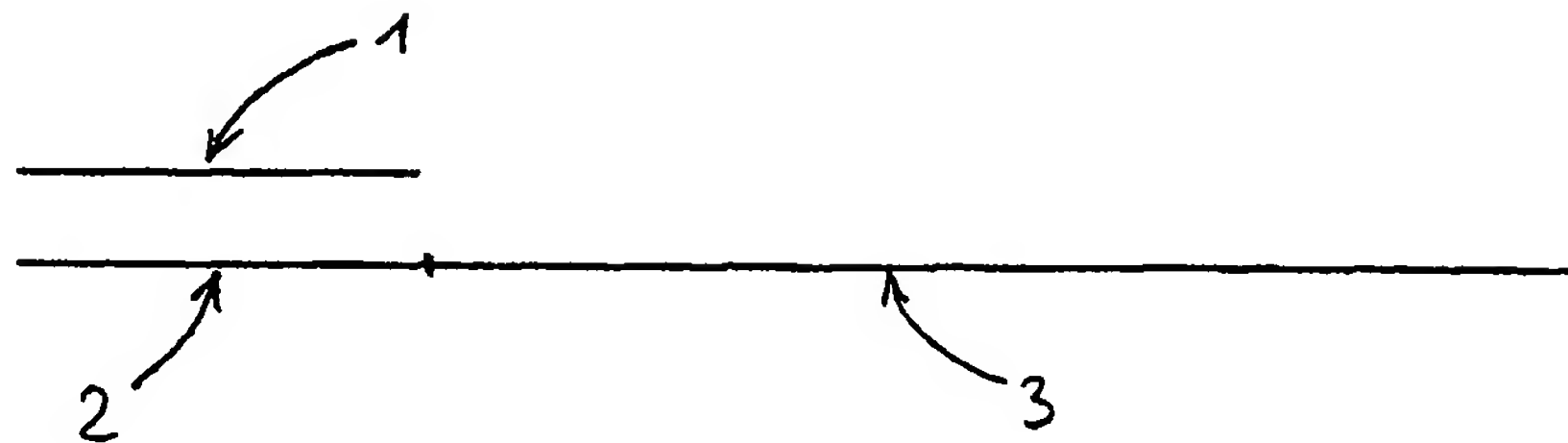
Mikroskop und Beleuchtung



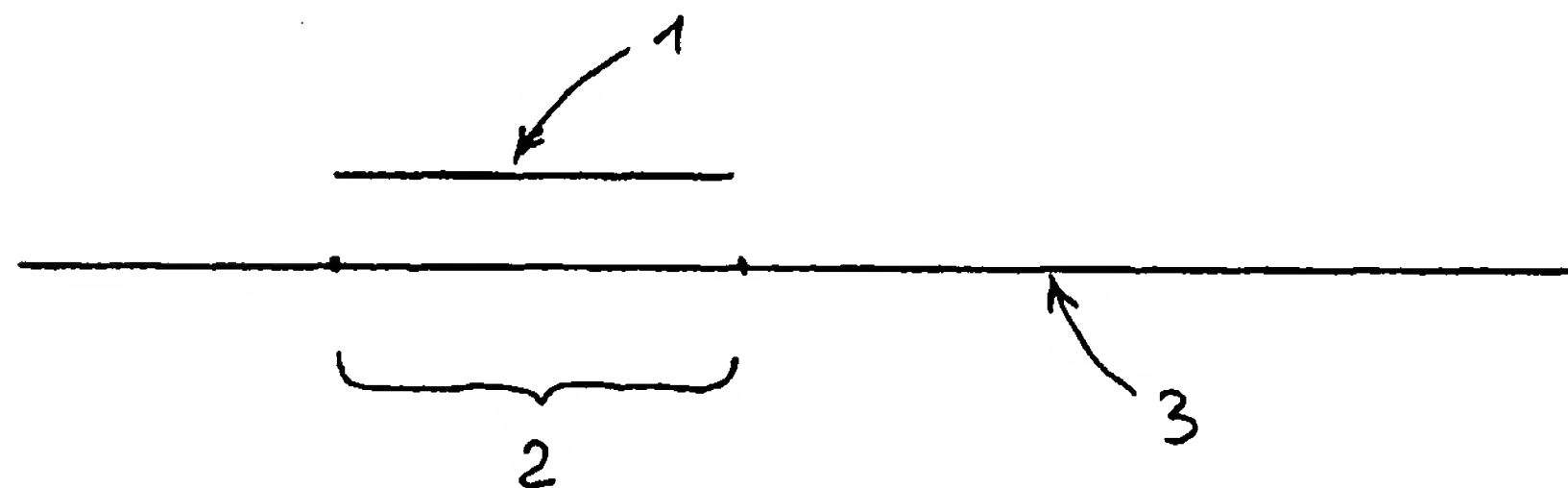


**Beispiel für allgemeine Struktur von NSKF-Primer-Komplexen**

2a)



2b)

**Legende:**

- 1) Primer
- 2) Primerbindungsstelle
- 3) NSKF

### Oligonukleotid-Konstruktion mit einer Primerbindungsstelle und blockierten Enden.

3a)

Oligonukleotidkomplex      5'-----3'  
                                      3'//-----//5'

Primerbindungsstelle      //-----//

3b)

Modifikation einer Seite des Oligonukleotides an beiden Strängen. Nur ein Ende steht für die Ligation zur Verfügung.

xxxxx-5'-----3'  
       kkk-3'//-----//5'

Die modifizierenden Gruppen "xxxxx-" und "kkk-" können z.B. zur Immobilisation oder Markierung dienen.

3c)

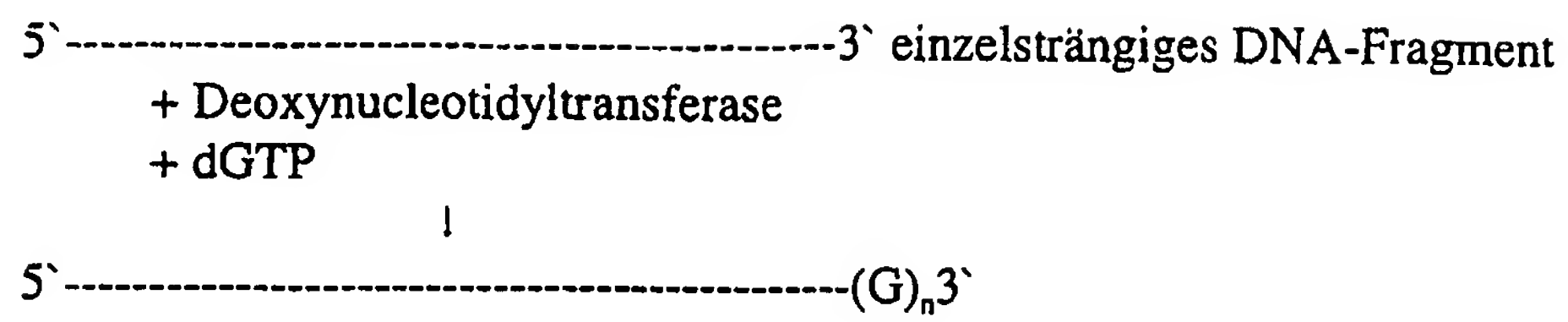
Resultat der Ligation des Oligonukleotides mit den DNA-Fragmenten

xxxxx-5'-----/-----//-----//3'-kkk  
       kkk-3//-----//-----/-----5'-xxxxx

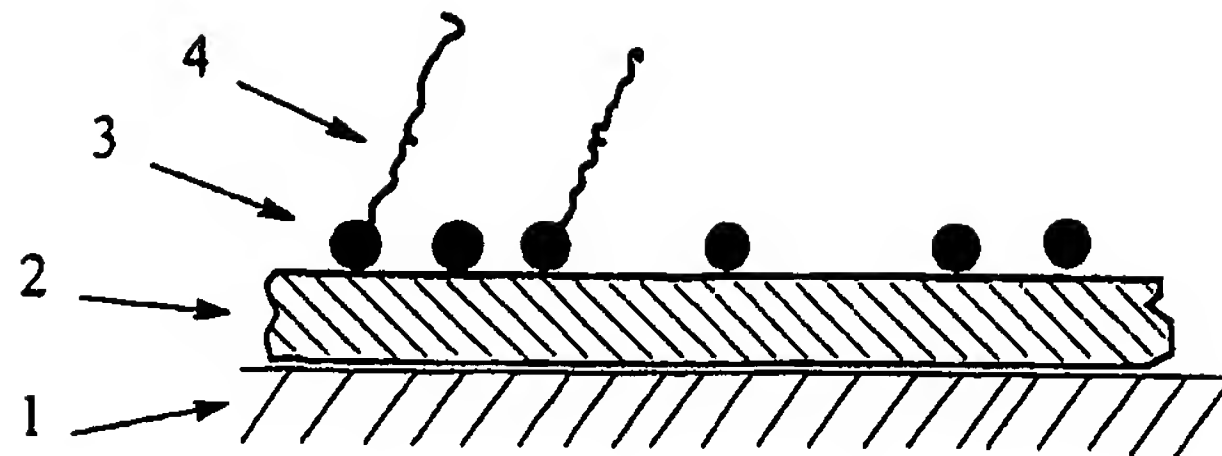
3d)

Konstruktion der DNA-Fragmente nach der Denaturierung

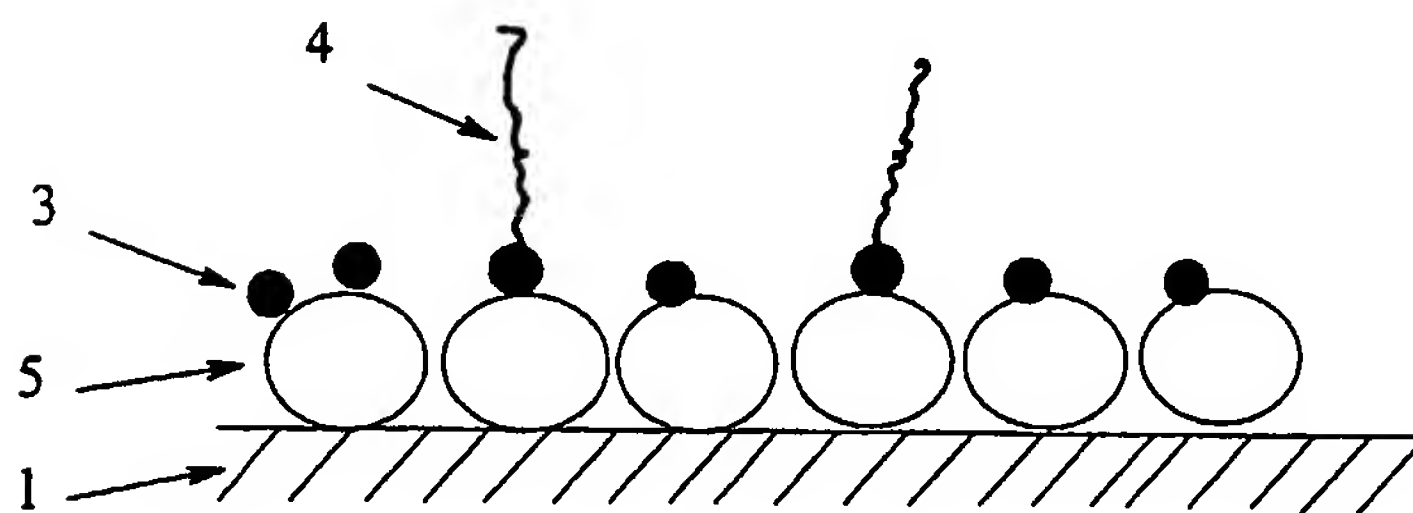
xxxxx-5'-----/-----//-----//3'-kkk

**Nukleotid-Tailing**

5a)



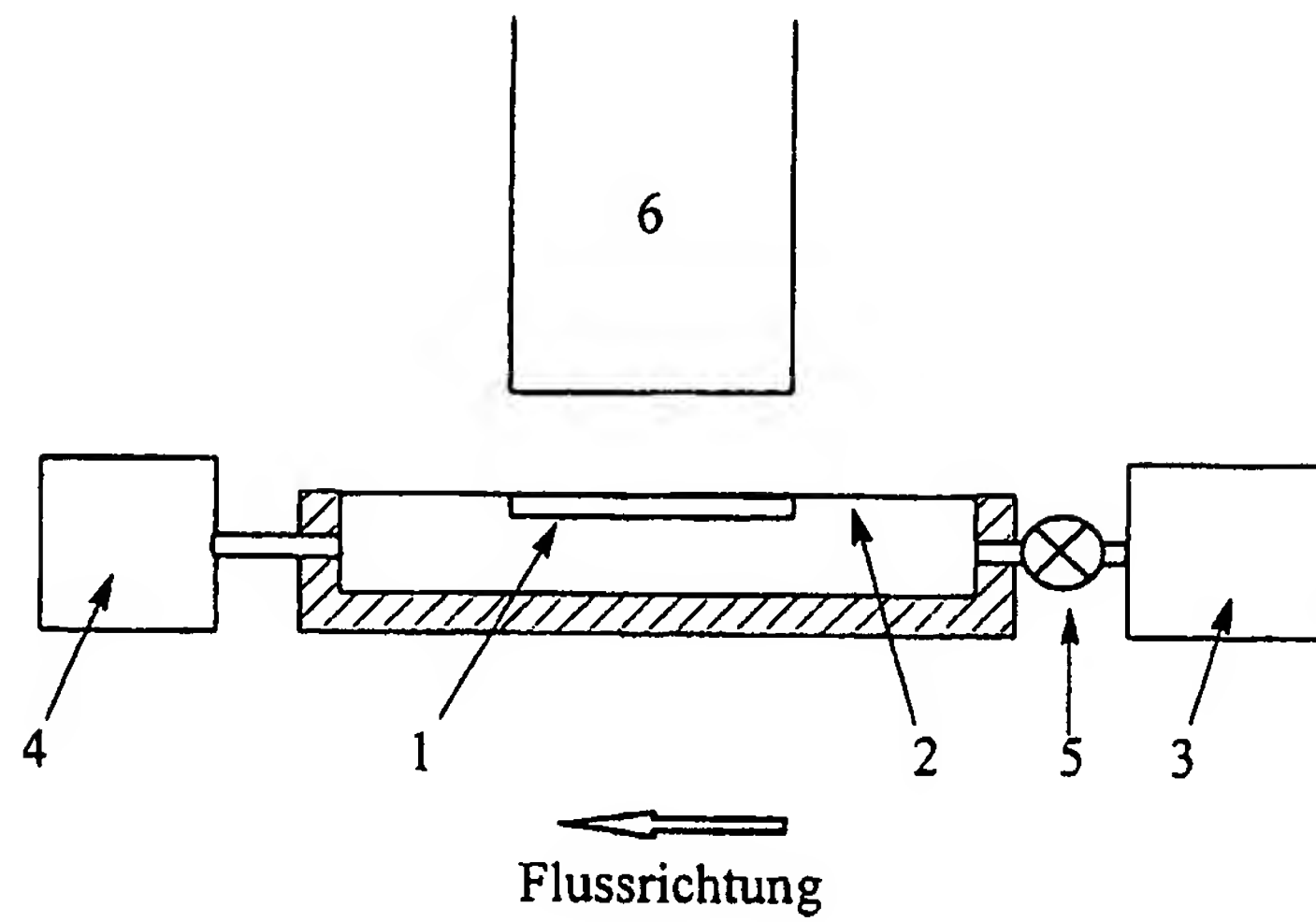
5b)

**Legende:**

- 1) feste Oberfläche
- 2) Gel
- 3) zur Immobilisation geeignete Gruppen, z.B. Avidin oder Streptavidin, oder Oligonukleotide
- 4) NSKFs
- 5) Agarosekügelchen

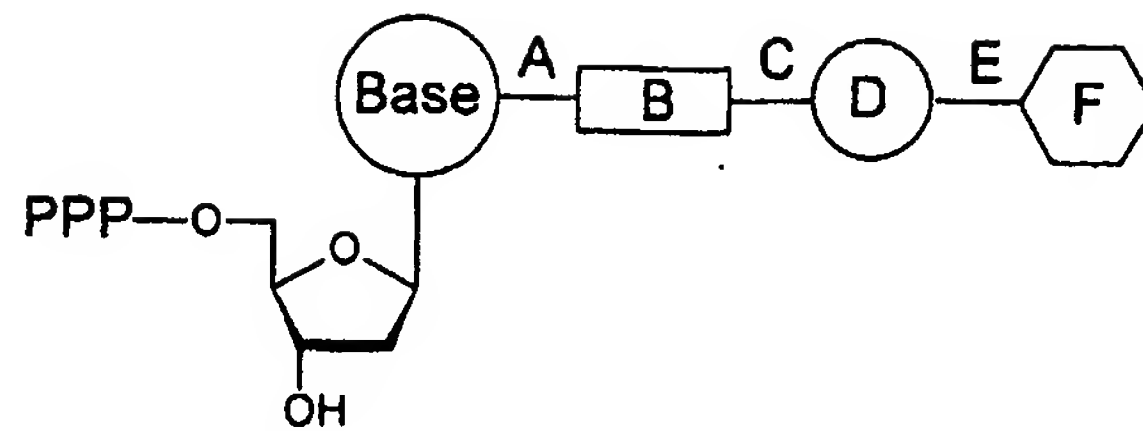
6/19

Fig. 6

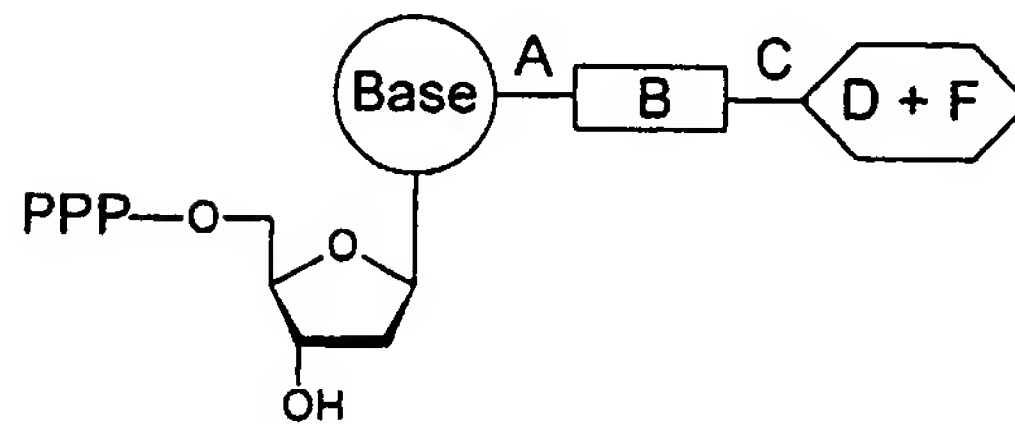
**Legende:**

- 1) Reaktionsoberfläche
- 2) feste Unterlage
- 3) Vorratsbehälter
- 4) Pumpe
- 5) Ventil
- 6) Objektiv des Mikroskops

7a)



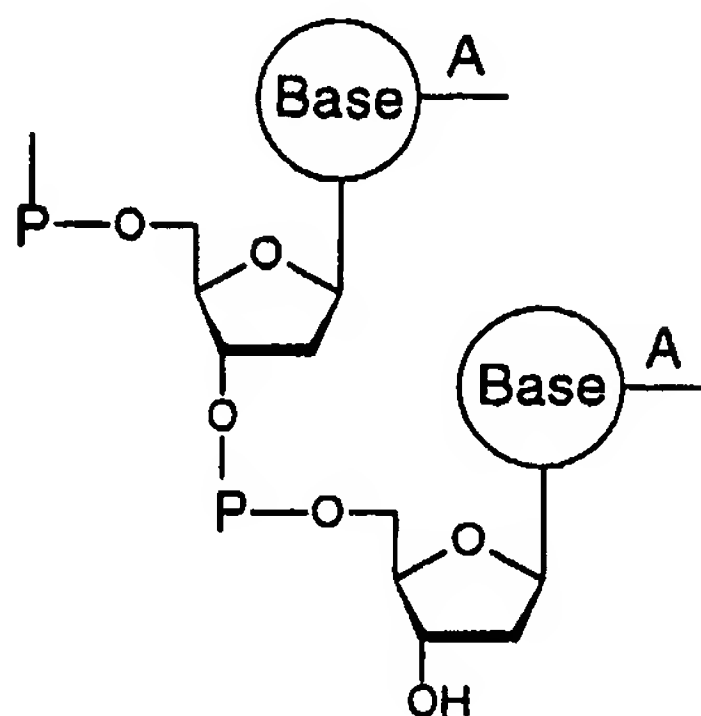
7b)

**Legende:**

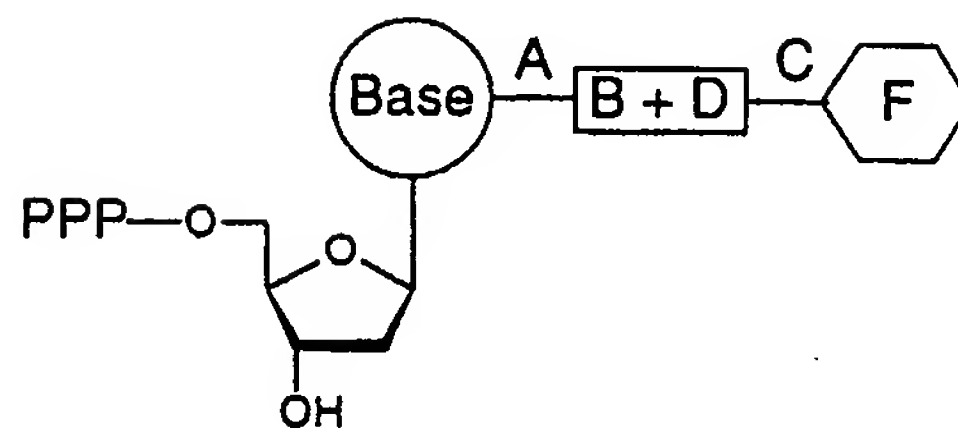
- A, C, E – Verbindungselemente im Linker
- A – Linkerrest nach der Spaltung
- B – spaltbare Verbindung / Gruppe
- D – zur Termination führende Gruppe
- F – Fluoreszenzfarbstoff (Marker)



7c) eingebaute modifizierte NTs\* nach der Spaltung



7d)



**Legende:**

A, C, E – Verbindungselemente im Linker

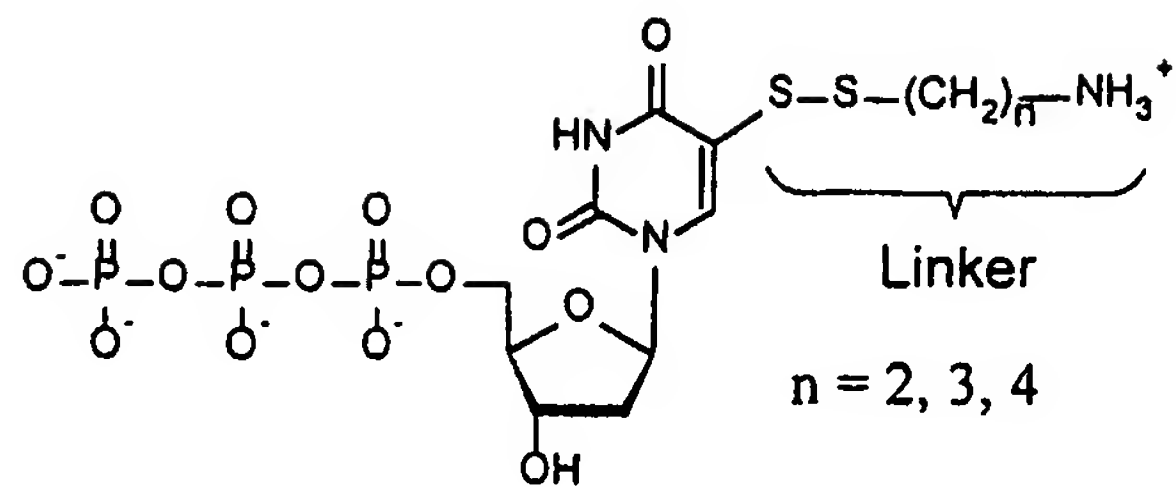
A – Linkerrest nach der Spaltung

B – spaltbare Verbindung / Gruppe

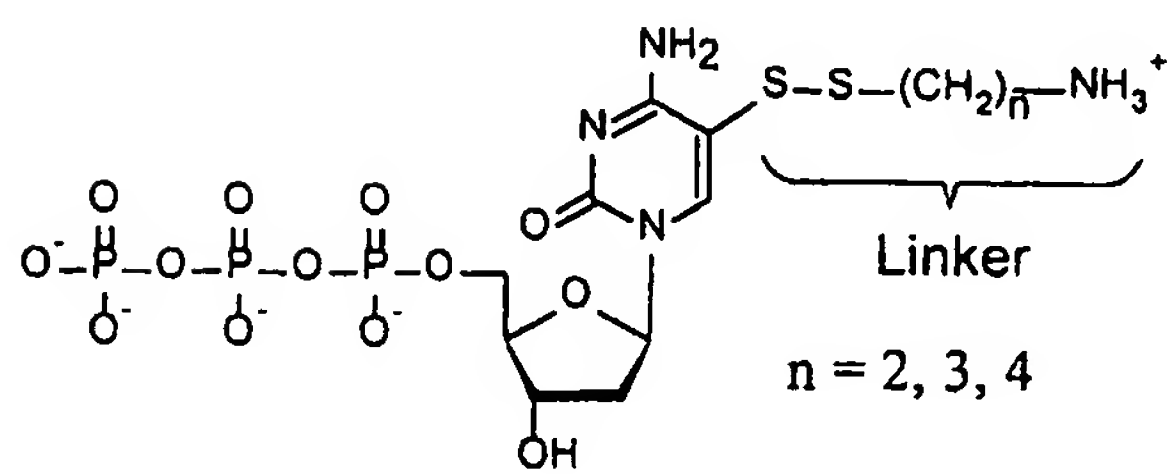
D – zur Termination führende Gruppe

F – Fluoreszenzfarbstoff (Marker)

7e-1)

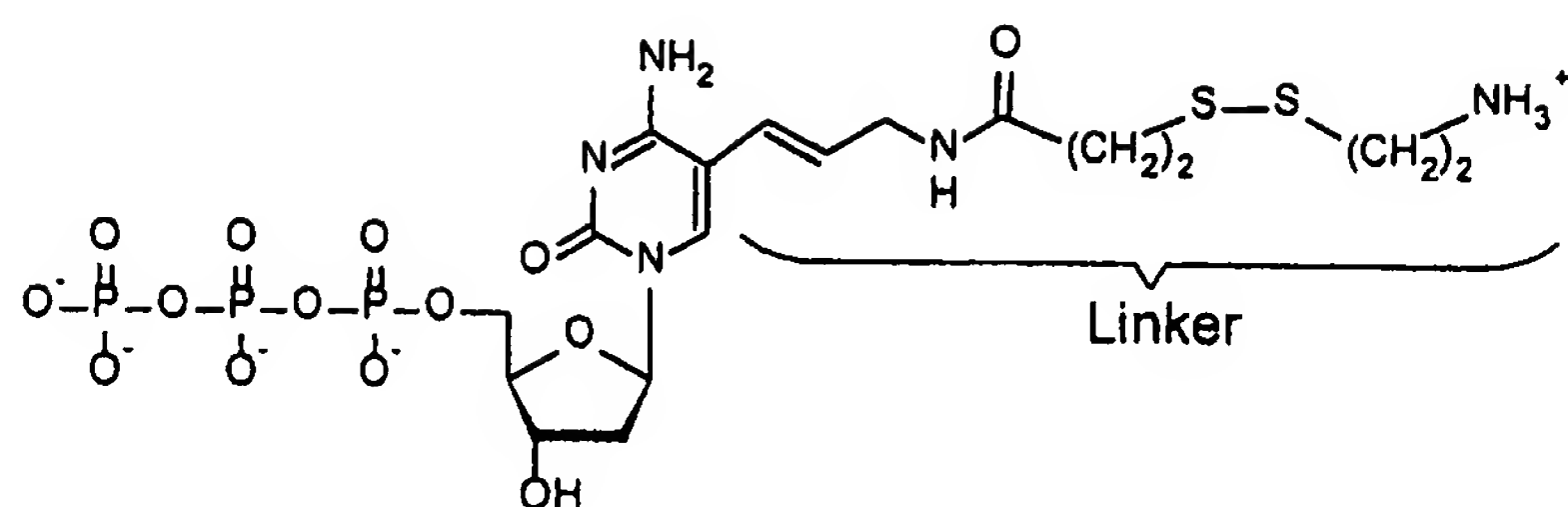


7e-2)

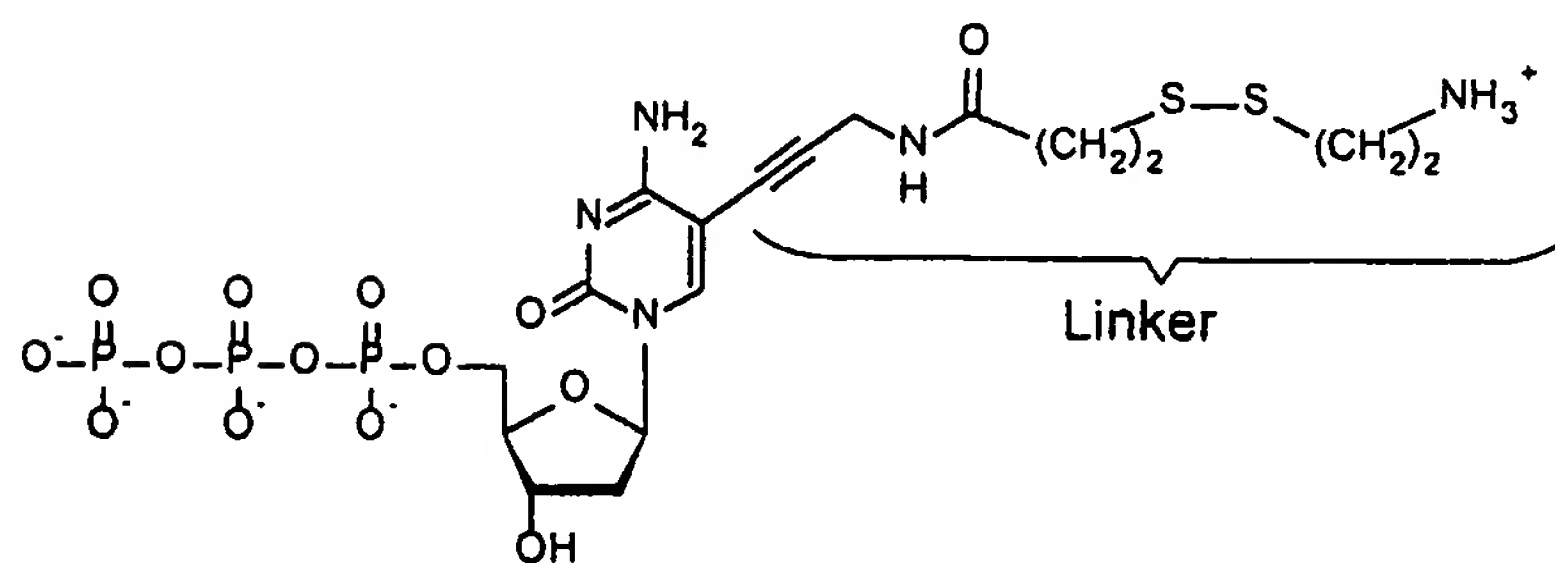




7f-3)



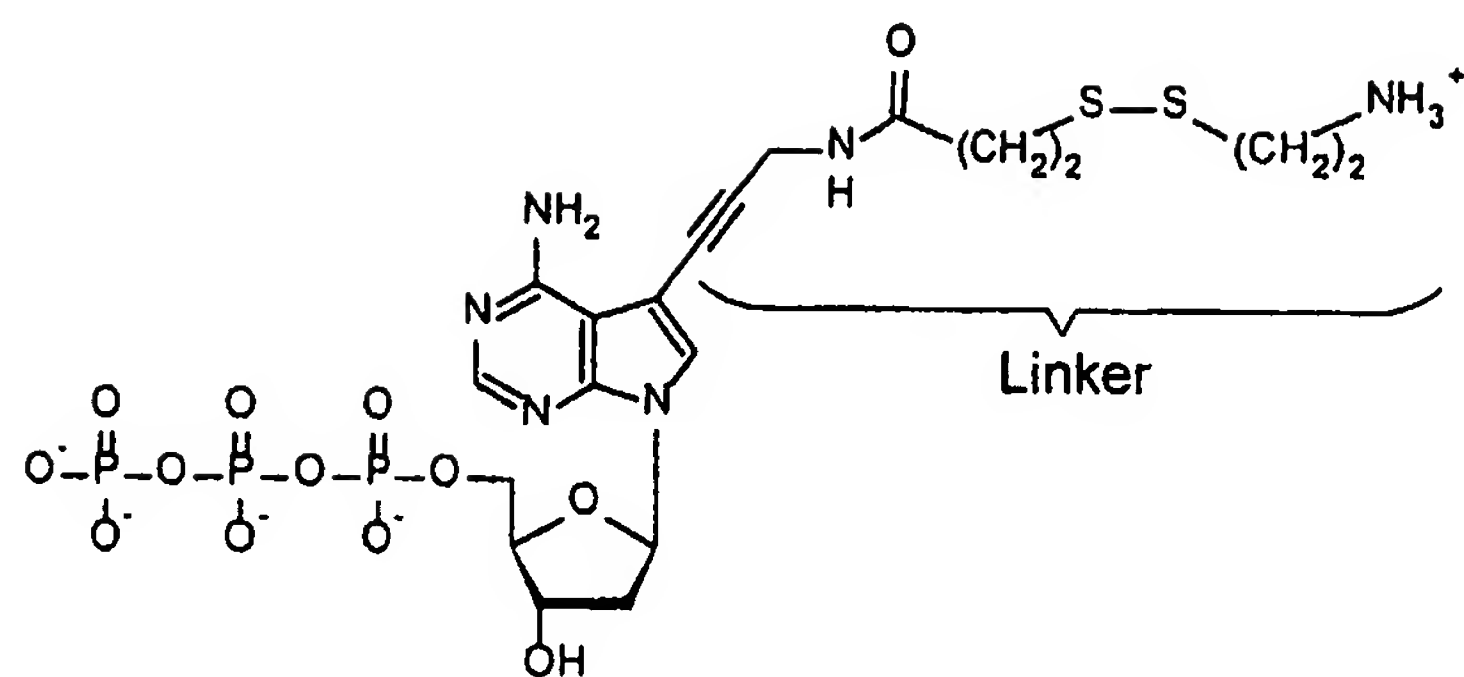
7f-4)



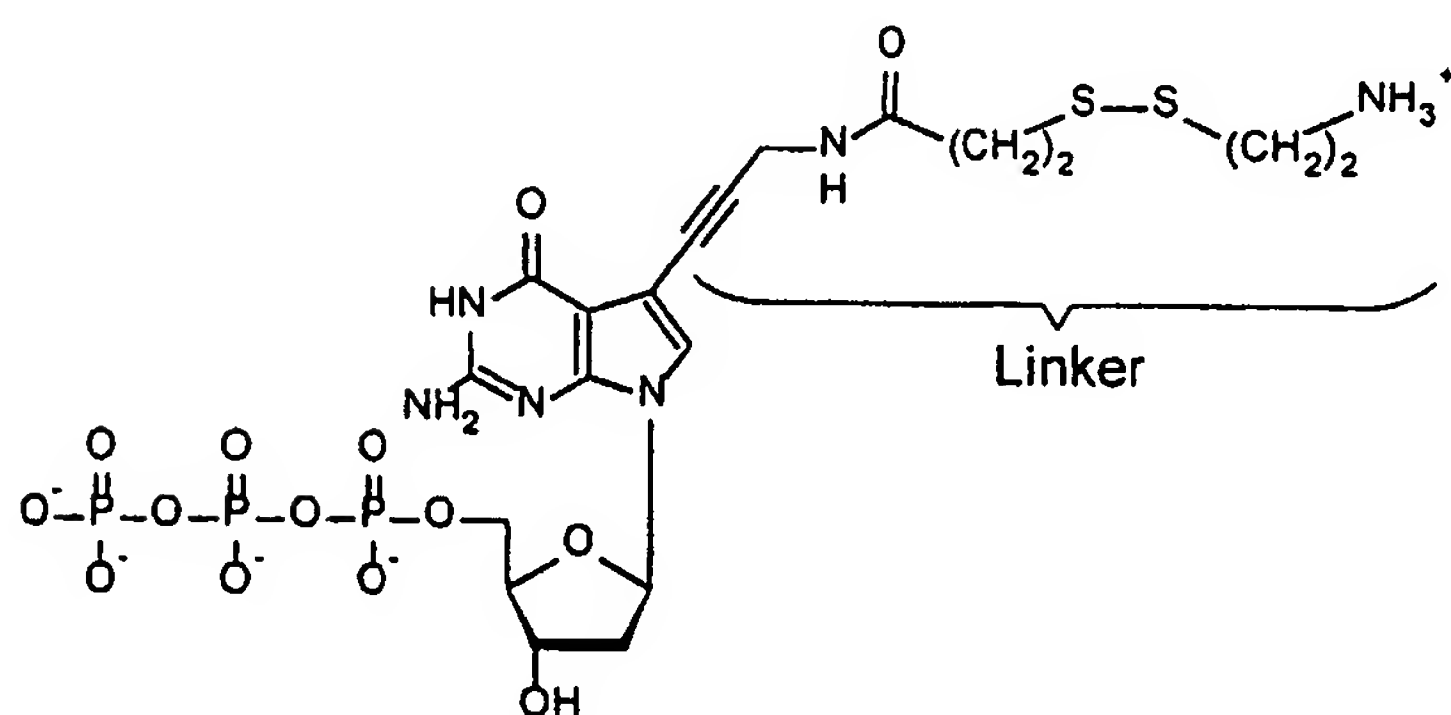
12/19

Fig. 7

7g-1)



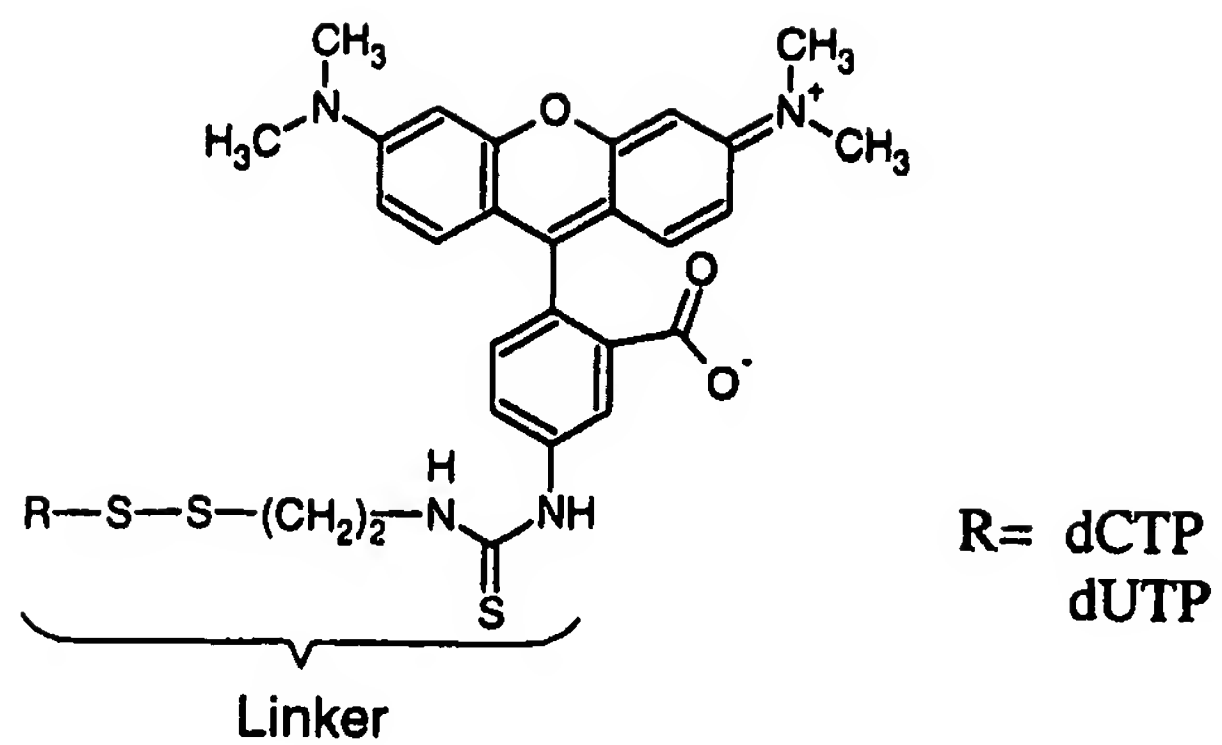
7g-2)



7h)

13/19

Fig. 7



7i)

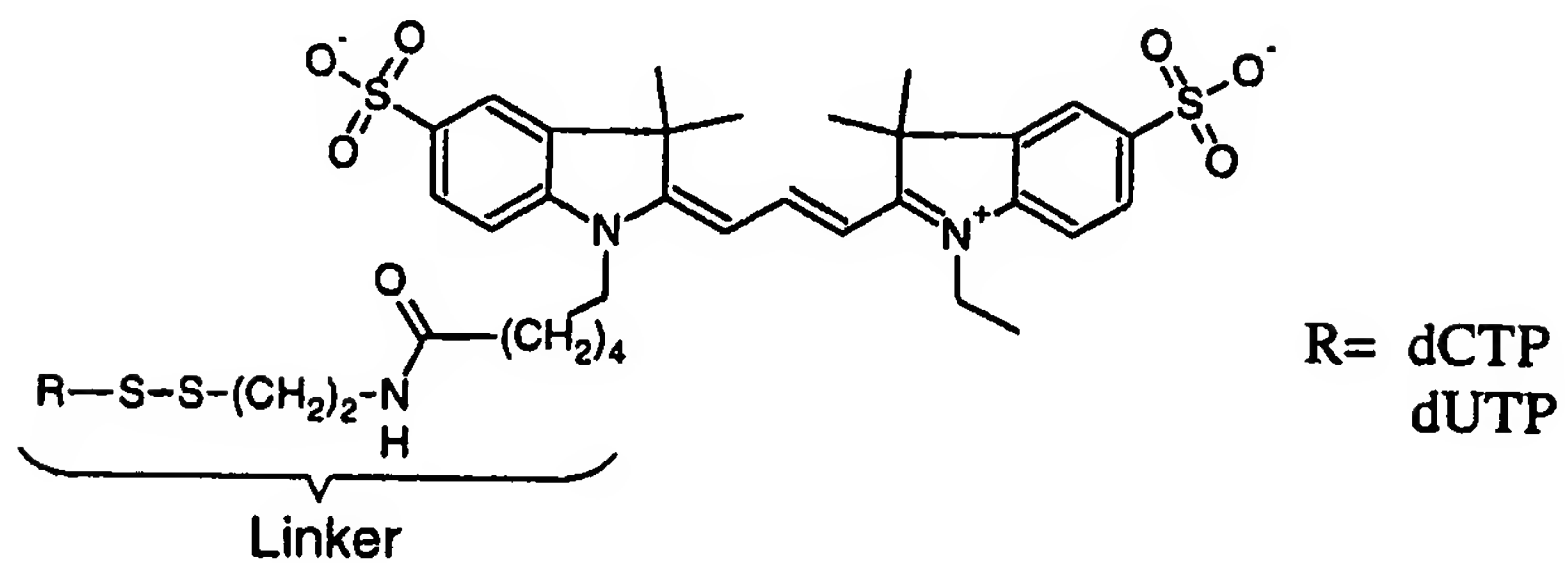
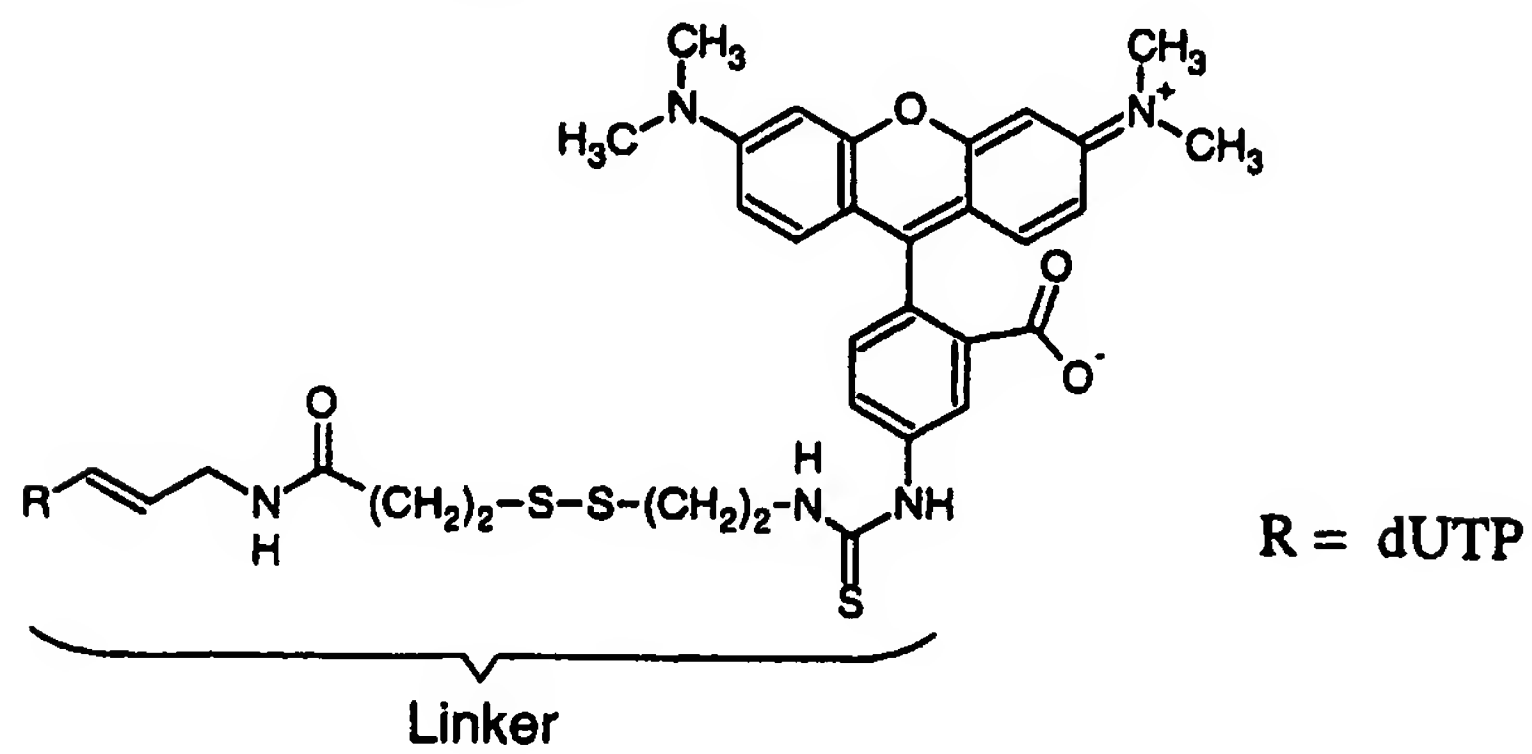
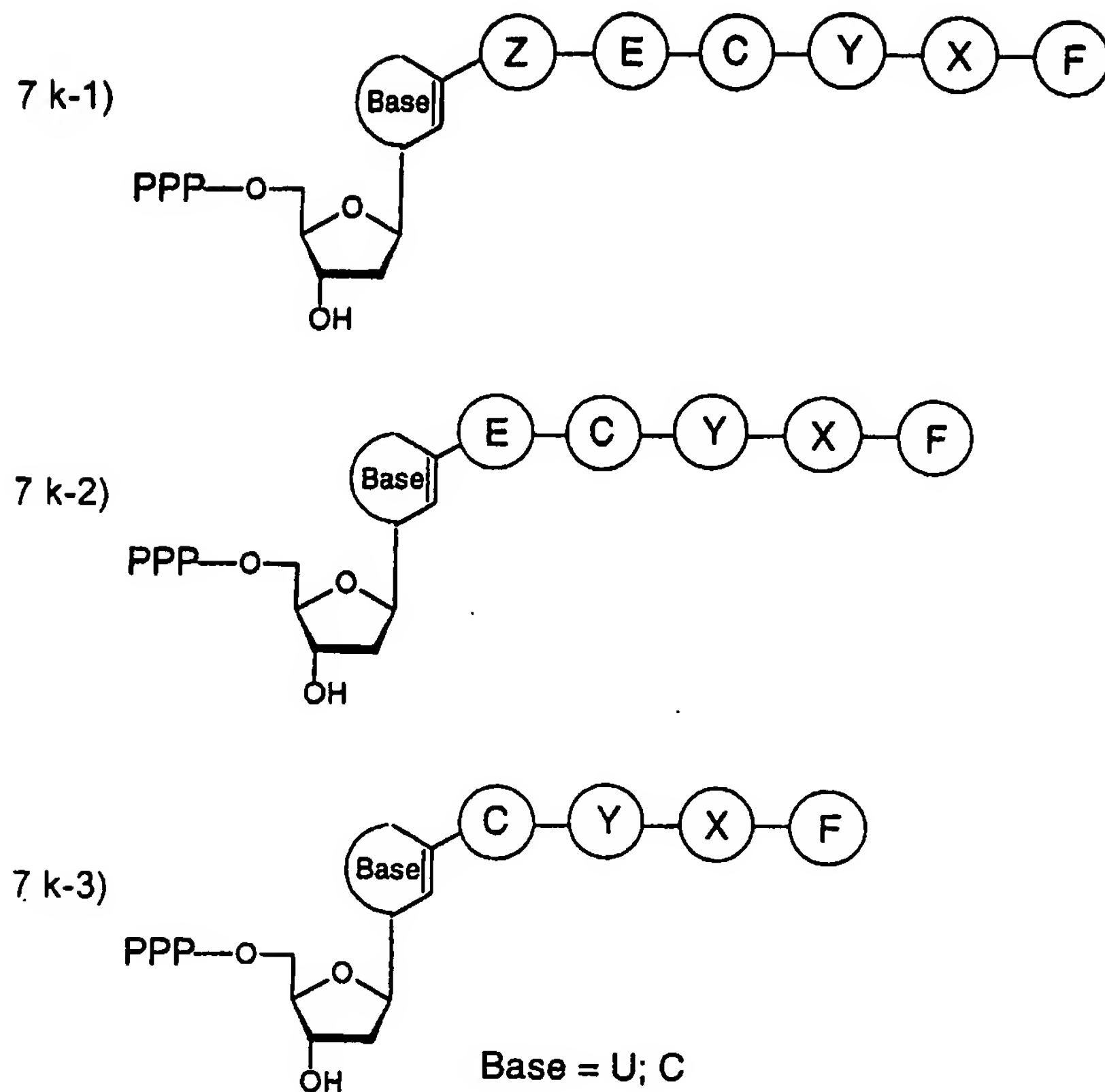
7j) dNTP-SS-TRITC (L 14)



Fig. 7



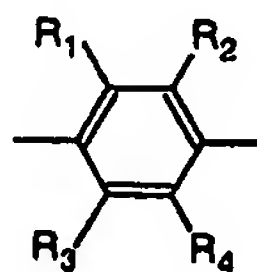
$R_1; R_2; R_3; R_4 = -OCH_3; -CH_3; -OH; -H; -NH_2; -NO_2; -O-COAlkyl; -Halogen$

$Z; X = -NH-; -S-; -O-; -CO-NH-; -CO-O-; -NH-CO-; -O-CO-$

$C = -CO-S-; -CO-O-; -S-S-; -S-CO-; -O-CO-$

$E = -(CH_2)_{n1}-; -(CH=CH-CH_2)_{n2}-; -(CH_2)_{n3}-NH-CO-(CH_2)_{n4}-$

$Y = -(CH_2)_{n1}-; -(CH=CH-CH_2)_{n2}-; -(CH_2)_{n3}-NH-CO-(CH_2)_{n4}-;$

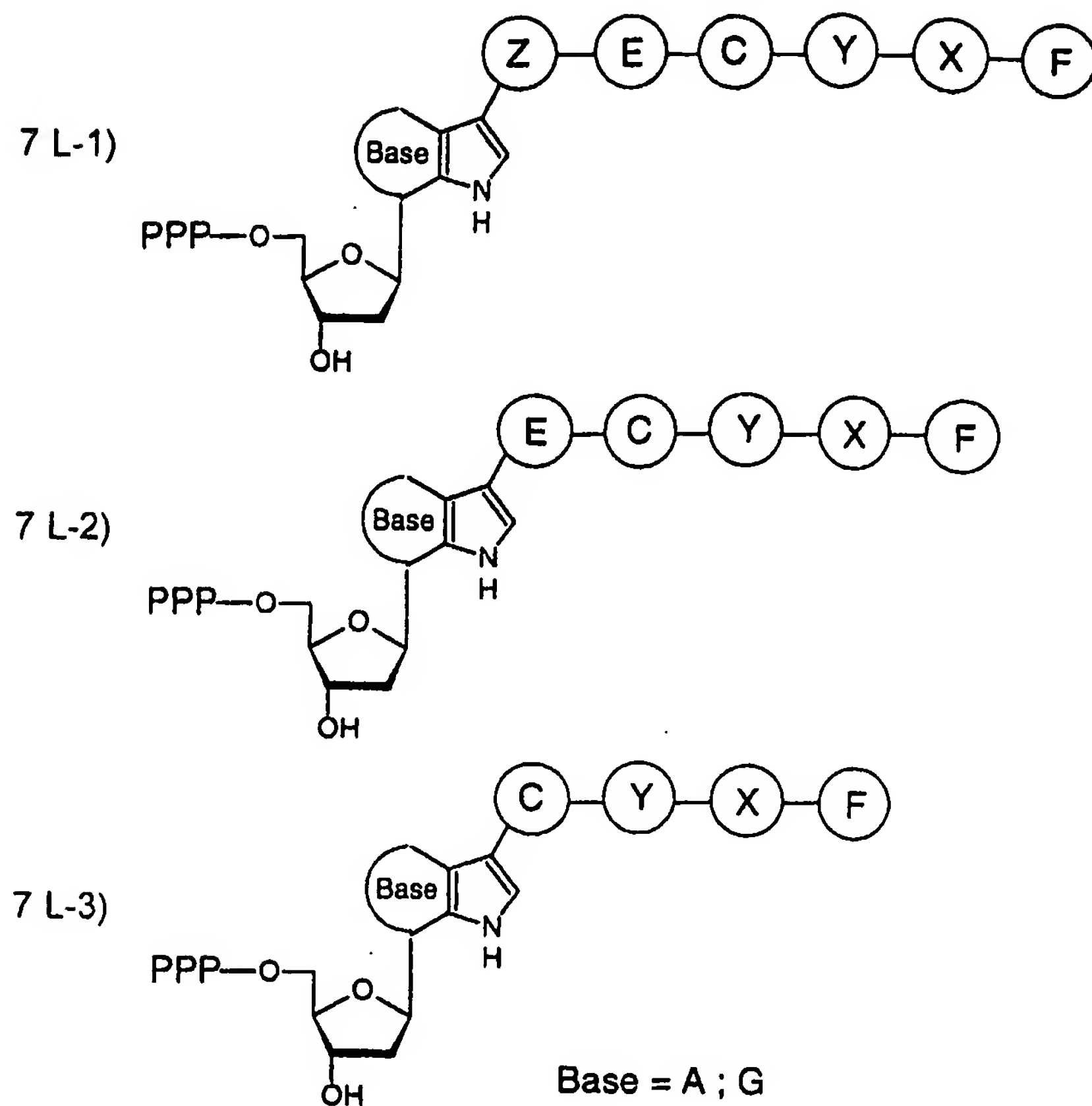


$n_1; n_2; n_3; n_4 = 1 - 5$

F = Fluoreszenzfarbstoff

15/19

Fig. 7

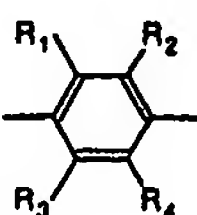


$R_1; R_2; R_3; R_4 = -OCH_3; -CH_3; -OH; -H; -NH_2; -NO_2; -O-COAlkyl; -Halogen$

$Z; X = -NH-; -S-; -O-; -CO-NH-; -CO-O-; -NH-CO-; -O-CO-$

$C = -CO-S-; -CO-O-; -S-S-; -S-CO-; -O-CO-$

$E = -(CH_2)_{n1}-; -(CH=CH-CH_2)_{n2}-; -(CH_2)_{n3}-NH-CO-(CH_2)_{n4}-$

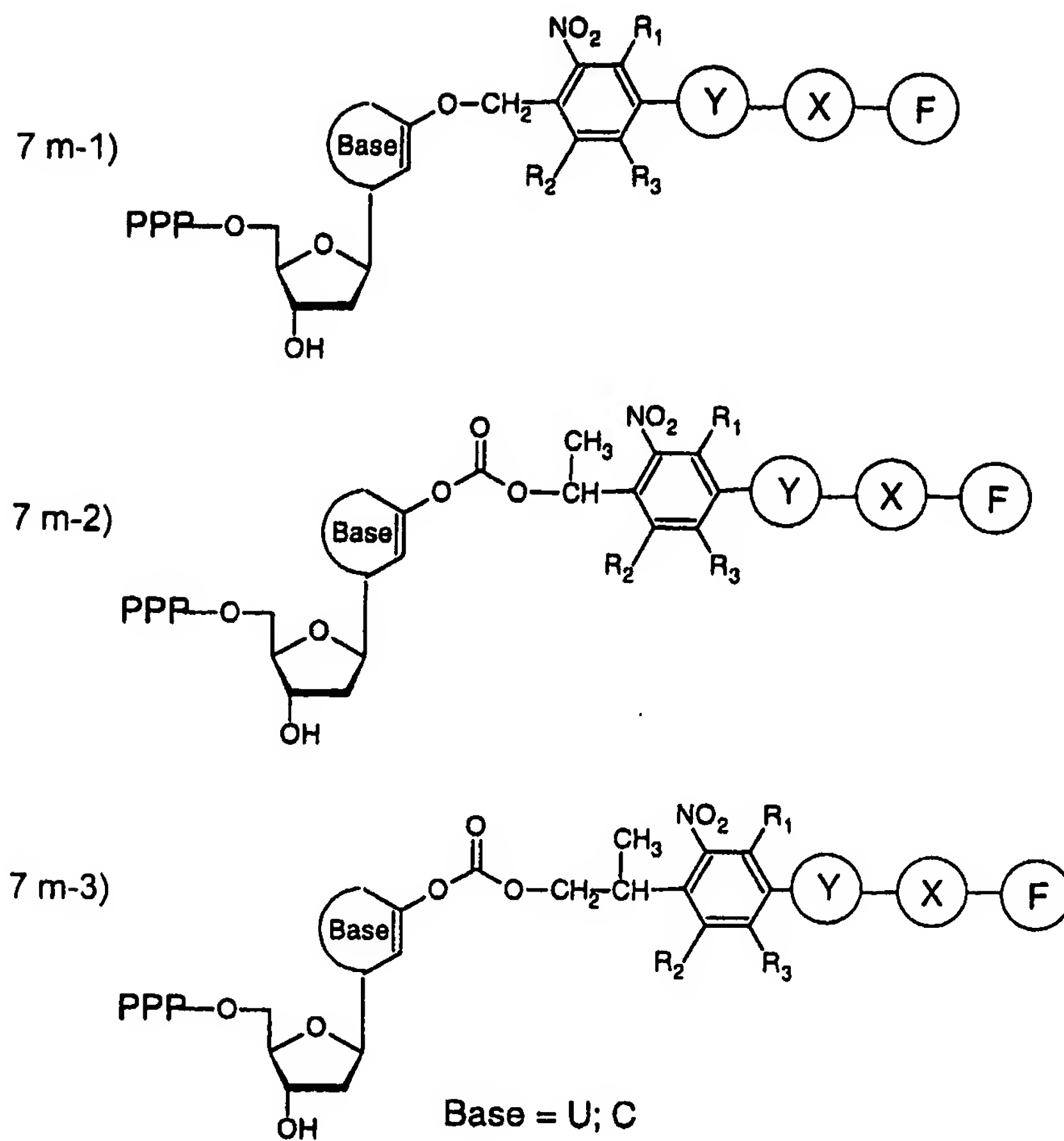
$Y = -(CH_2)_{n1}-; -(CH=CH-CH_2)_{n2}-; -(CH_2)_{n3}-NH-CO-(CH_2)_{n4}-;$  

$n_1; n_2; n_3; n_4 = 1 - 5$

F = Fluoreszenzfarbstoff

16/19

Fig. 7



$R_1, R_2, R_3, R_4 = -OCH_3; -CH_3; -OH; -H; -NH_2; -NO_2; -O-COAlkyl; -Halogen$

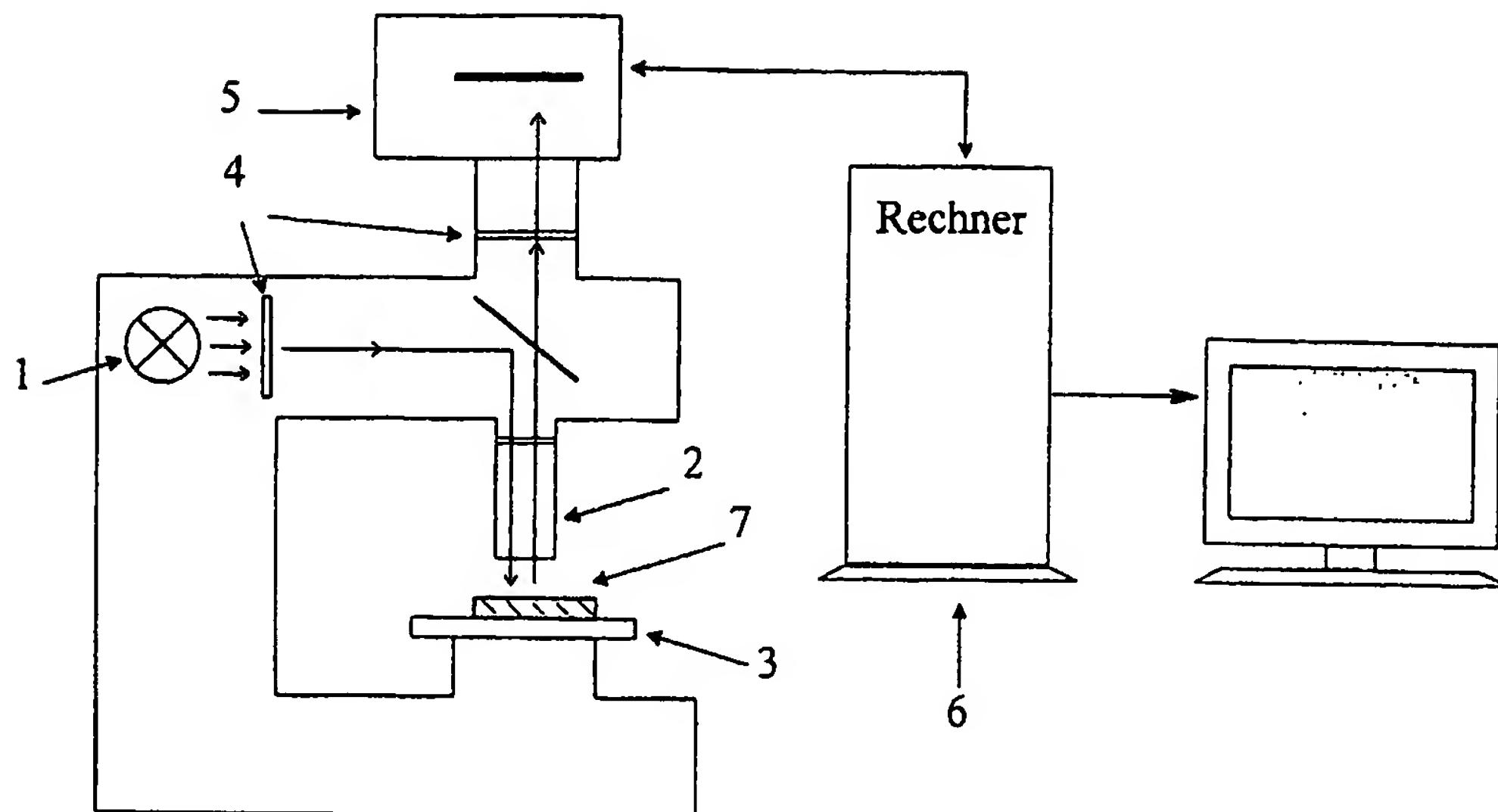
$Z; X = -NH-; -S-; -O-; -CO-NH-; -CO-O-; -NH-CO-; -O-CO-$

$C = -CO-S-; -CO-O-; -S-S-; -S-CO-; -O-CO-$

$E; Y = -(CH_2)_{n_1}-; -(CH=CH-CH_2)_{n_2}-; -(CH_2)_{n_3}-NH-CO-(CH_2)_{n_4}-$

$n_1; n_2; n_3; n_4 = 1 - 5$

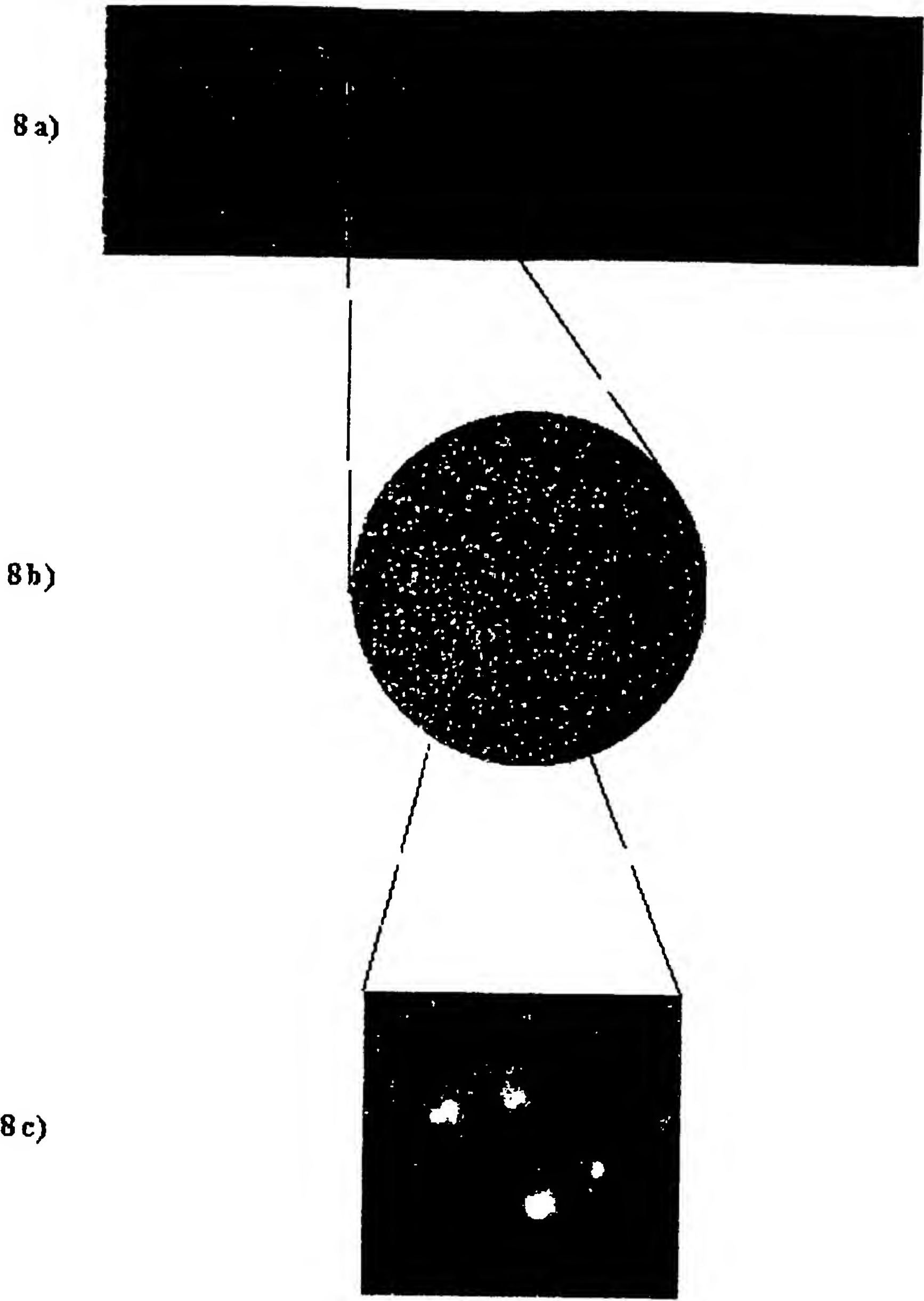
F = Fluoreszenzfarbstoff



**Legende:**

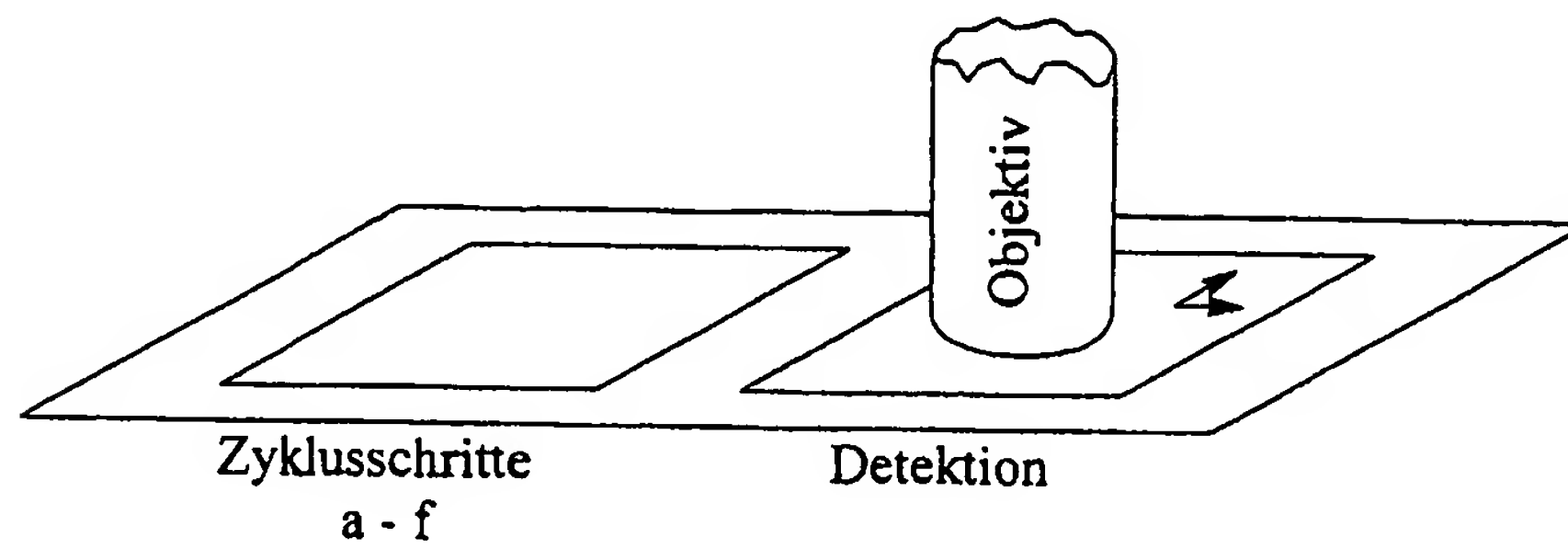
- 1) Lichtquelle zur Anregung der Fluoreszenz
- 2) Lichtleitender Teil
- 3) Scanntisch
- 4) Vorrichtung zur Selektion von Spektren
- 5) Detektionsvorrichtung
- 6) Computer mit Steuerungs- und Analysefunktionen
- 7) Oberfläche mit immobilisierten NSKFs

Fig. 8



Sequenzierungsreaktion auf 2 getrennten Oberflächen.

Während eine Oberfläche gescannt wird, werden auf der andern Oberfläche alle anderen Zyklusschritte durchgeführt.





## SEQUENZPROTOKOLL

<110> Dmitri Tcherkassov

<110> Prof.Dr. Walther Traut

<110> Dr. Englbert Bäuml

<110> Christian Hennig

<110> Norman Kock

<120> Verfahren zur Analyse von Nukleinsäureketten

<130> P060086

<140>

<141>

<160> 9

<170> PatentIn Ver. 2.1

<210> 1

<211> 19

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primer (T,-19-Primer)

<400> 1

taatacgact cactatagg

19

<210> 2

<211> 19

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Primerbindungsstelle

<400> 2

attatgctga gtgatatcc

19

<210> 3

<211> 12

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Sequenz aus Beispiel 4

<400> 3

accaaaacac cc

12

<210> 4

<211> 32

<212> DNA

<213> Künstliche Sequenz

<220>

<223> Beschreibung der künstlichen Sequenz: Sequenz aus Beispiel 4

<400> 4  
atcatcgttc gaaatatcga tcgcctgatg cc 32

<210> 5  
<211> 24  
<212> DNA  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz: Sequenz aus Beispiel 4

<400> 5  
atcgttcgaa atatcgatcg cctg 24

<210> 6  
<211> 24  
<212> DNA  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz: Sequenz aus Beispiel 4

<400> 6  
attgttcgaa atatcgatcg cctg 24

<210> 7  
<211> 24  
<212> DNA  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz: Sequenz aus Beispiel 4

<400> 7  
tagcaagctt tatagctagc ggac 24

<210> 8  
<211> 24  
<212> DNA  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz: Sequenz aus Beispiel 4

<400> 8  
taacaagctt tatagctagc ggac 24

<210> 9  
<211> 11  
<212> DNA  
<213> Künstliche Sequenz

<220>  
<223> Beschreibung der künstlichen Sequenz: Sequenz aus Beispiel 4

<400> 9  
acaacaacac c 11

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**